

REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: KP-7501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 108..1682
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..3327
FEATURE:
NAME/KEY: CDS
LOCATION: 3438..4442
FEATURE:
NAME/KEY: CDS
LOCATION: 5458..7107
FEATURE:
NAME/KEY: CDS
LOCATION: 7271..9121
FEATURE:
NAME/KEY: CDS
LOCATION: 9234..15782
US-08-348-891A-1

Query Match 70.4%; Score 17.6; DB 1; Length 15894;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 gatctgctgttgaagcgacag 25
|||||
DB 12940 GATCTGCTGTGAGATGACACAG 12963

RESULT 9
US-08-905-817-1
Sequence 1, Application US/08905817
Patent No. 5824777
GENERAL INFORMATION:
APPLICANT: SASAKI, Keiko
APPLICANT: MORI, Takayuki
APPLICANT: MAKINO, Satoshi
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,817
FILING DATE: 04-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,891

FILING DATE: 25-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,400
FILING DATE: 10-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-293625
FILING DATE: 14-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: KP-7501A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 108..1682
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..3327
FEATURE:
NAME/KEY: CDS
LOCATION: 3438..4442
FEATURE:
NAME/KEY: CDS
LOCATION: 5458..7107
FEATURE:
NAME/KEY: CDS
LOCATION: 7271..9121
FEATURE:
NAME/KEY: CDS
LOCATION: 9234..15782
US-08-905-817-1

Query Match 70.4%; Score 17.6; DB 1; Length 15894;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacag 25
|||||
DB 12940 GATCTGCTGTGAGATGACACAG 12963

RESULT 10
US-08-885-469-1/C
Sequence 1, Application US/08885469
Patent No. 6280739
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Racie, Lisa A.
APPLICANT: Lavalley, Edward R.
APPLICANT: Werberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael
TITLE OF INVENTION: SECRETED PROTEINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,574C
FILING DATE: 18-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/282,012
FILING DATE: 29-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/097,997
FILING DATE: 29-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/118,968
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0656,0370002/SLF/LBB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3561
US-08-665-574C-12

Query Match 70.4%; Score 17.6; DB 3; Length 3561;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacagc 25
||||| ||||| ||||| |||||
DB 3021 GATCTGCTGCGCGACAGCAGCAGCAG 2998

RESULT 7
US-08-946-994-12/C
Sequence 12, Application US/08946994
Patent No. 6210654
GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Witthuhn, Bruce A.
APPLICANT: Queller, Frederick W.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytochrome Signal
TITLE OF INVENTION: Transduction
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,994
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/665,574

FILING DATE: 18-JUN-1996
APPLICATION NUMBER: 08/282,012
FILING DATE: 29-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/097,997
FILING DATE: 29-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/118,968
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656,0370002/SLF/GKT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3561
US-08-946-994-12

Query Match 70.4%; Score 17.6; DB 4; Length 3561;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacagc 25
||||| ||||| ||||| |||||
DB 3021 GATCTGCTGCGCGACAGCAGCAGCAG 2998

RESULT 8
US-08-348-891A-1
Sequence 1, Application US/08348891A
Patent No. 5654136
GENERAL INFORMATION:
APPLICANT: SASAKI, Keiko
APPLICANT: MORI, Takayuki
APPLICANT: MAKINO, Satoshi
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,891A
FILING DATE: 25-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,400
FILING DATE: 10-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-293625
FILING DATE: 14-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=1154
OTHER INFORMATION: /EC_number=2.3.1.47
OTHER INFORMATION: /product="KAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioF"
OTHER INFORMATION: /number=2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3043
OTHER INFORMATION: /EC_number=6.3.3.3
OTHER INFORMATION: /product="DTB synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioD"
OTHER INFORMATION: /number=4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name="bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name="bioD RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtctgctgtttgaagcgagcag 25
|||||
Db 179 GGATCTGCTGTGGAAGCGAGCAG 203

RESULT 5
US-08-097-997A-12/C
Sequence 12, Application US/08097997A
Patent No. 5728536

GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wilthuhn, Bruce A.
APPLICANT: Ouellet, Frederick W.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal
TITLE OF INVENTION: Transduction
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/097,997A
FILING DATE: 29-JULY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656.0370000/SLF/GKT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3561
US-08-097-997A-12

Query Match 70.4%; Score 17.6; DB 1; Length 3561;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgtttgaagcgagcag 25
|||||
Db 3021 GATCTGCTGCGGAGAGCAGCAG 2998

RESULT 6
US-08-665-574C-12/C
Sequence 12, Application US/08665574C
Patent No. 6136595
GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wilthuhn, Bruce A.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine
TITLE OF INVENTION: Signal Transduction
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: /aminoacids="aminoacids"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgtcgttgaagcgcagcag 25
DB 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1

```

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcag 25
|
Db 63 GGATCTGCTGTTGAAGCGAGCAG 87

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcag 25
|
Db 63 GGATCTGCTGTTGAAGCGAGCAG 87

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:23:04 : Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25

Sequence: 1 ggaatcgctgttgaagcagcagcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.6	70.4	3561	1	US-08-097-997A-12
6	17.6	70.4	3561	3	US-08-665-574C-12
7	17.6	70.4	3561	4	US-08-946-994-12
8	17.6	70.4	15894	4	US-08-348-891A-1
9	17.6	70.4	15894	4	US-08-905-817-1
10	16.6	66.4	1835	1	US-08-885-469-1
11	16.6	66.4	1835	1	US-08-361-611-3
12	16.6	66.4	1835	1	US-08-565-655-3
13	16.6	66.4	1835	2	US-08-946-967-3
14	16.2	64.8	1694	1	US-08-698-551-9
15	16.2	64.8	1694	2	US-08-602-228-9
16	16.2	64.8	1694	2	US-08-494-440B-9
17	16.2	64.8	1694	2	US-08-533-901B-9
18	16.2	64.8	1694	2	US-08-839-032A-9
19	16.2	64.8	1694	2	US-08-839-032A-9
20	16.2	64.8	1694	2	US-08-839-032A-9
21	16.2	64.8	1810	1	US-07-755-573C-7
22	16.2	64.8	2473	1	US-08-698-551-17
23	16.2	64.8	2473	2	US-08-839-032A-17
24	16.2	64.8	1080	1	US-08-225-757B-1
25	16.2	64.8	1080	2	US-08-722-050-1
26	16.2	64.8	1248	3	US-08-910-505-3
27	16.2	64.8	1251	3	US-08-910-505-3

Thu Dec 27 16:12:19 2001

us-09-396-196f-4.rng

Page 24

Search completed: December 26, 2001, 12:18:29
Job time: 6870 sec

CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.

XX Sequence 2526 BP; 696 A; 495 C; 597 G; 738 T; 0 other;

Query Match 68.0%; Score 17; DB 20; Length 2526;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgtttgaagcagcag 25
||| ||||| || ||||| |||||
Db 331 ggaagctgctattggaagcagcag 355

RESULT 44

AAA28691
ID AAA28691 standard; DNA: 2526 BP.

AC AAA28691;

DT 29-AUG-2000 (first entry)

DE C. pneumoniae CPN100397 open reading frame.

XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KW anti-arteriosclerotic; vaccine; ds.

XX Chlamydia pneumoniae.

PN WO200024765-A2.

PD 04-MAY-2000.

PF 28-OCT-1999; 99WO-CA00992.

PR 28-OCT-1998; 98US-0106034.

PR 28-OCT-1998; 98US-0106039.

PR 28-OCT-1998; 98US-0106042.

PR 28-OCT-1998; 98US-0106044.

PR 29-OCT-1998; 98US-0106072.

PR 29-OCT-1998; 98US-0106073.

PR 29-OCT-1998; 98US-0106074.

PR 29-OCT-1998; 98US-0106087.

PR 02-NOV-1998; 98US-0106587.

PR 02-NOV-1998; 98US-0107034.

PR 02-NOV-1998; 98US-0107035.

PA (CONN-) CONNAUGHT LAB LTD.

PI Murdia AD, Oomen RP, Wang J;

DR WPI: 2000-350688/30.

DR P-PSDB; AAY92818.

PT Chlamydia antigens and the proteins they encode, useful for

PT vaccinating against Chlamydia infections that affect the respiratory

PT tract

PS Claim 2; Fig 1; 226pp; English.

XX The nucleic acids may be used for the recombinant production of the

CC Chlamydia polypeptides (either in vivo or in vitro) according to standard

CC recombinant DNA methodologies. The polypeptides may then be used to

CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as

CC C. pneumoniae, are pathogens responsible for upper respiratory tract

CC infections such as community acquired pneumonia, acute respiratory

CC disease and bronchitis and may be implicated in atherosclerotic changes

CC and asthma. The nucleic acids may also be used as probes for detecting

CC the presence of Chlamydia nucleic acids in samples (and therefore

CC diagnose infections) and the proteins may be used as antigens for the

CC production of antibodies that may be used to detect Chlamydia proteins

CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).
XX Sequence 2526 BP; 693 A; 493 C; 601 G; 739 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2526;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgtttgaagcagcag 25
||| ||||| || ||||| |||||
Db 331 ggaagctgctattggaagcagcag 355

RESULT 45

AAA61153/c
ID AAA61153 standard; DNA: 2612 BP.

AC AAA61153;

DT 12-OCT-2000 (first entry)

DE SEN virus genome fragment SEQ ID NO: 94.

XX SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;

XX proliferative disorder; hepatopathy; hepatitis; viral infection;

XX vaccination; gene therapy; ds.

XX Hepatitis virus.

PN WO200028039-A2.

PD 18-MAY-2000.

PF 09-NOV-1999; 99WO-EP08566.

PR 10-NOV-1998; 98IT-MI02437.

PR 30-APR-1999; 99IT-MI00923.

PR 14-MAY-1999; 99EP-0830298.

PR 16-JUL-1999; 99EP-0113932.

PA (DIAS-) DIASORIN SRL.

PI Primi D, Fioralisi G, Mantero GL, Mattioli S, Sottini A;

PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;

DR WPI: 2000-376551/32.

PT Nucleic acids representing the genome of the SEN virus (SENV) and

PT encoded proteins, useful for treatment of hepatopathies, inflammatory

PT diseases and proliferative disorders such as cancer -

PS Claim 1; Page 344-345; 392pp; English.

XX The present invention is concerned with the sequence of the genome of the

XX SEN virus (SENV), and the proteins encoded by it. SENV is thought to be

XX the cause of hepatopathies which are not linked to the presence of the

XX hepatitis A, B and E viruses in man. The genome and proteins of this

XX virus can be used in gene therapy and vaccination against the virus,

XX which also causes disorders of the gastrointestinal tract, including

XX Crohn's disease and lupus erythematosus, inflammatory diseases, and

XX proliferative disorders such as cancer.

XX Sequence 2612 BP; 828 A; 736 C; 545 G; 503 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2612;

Best Local Similarity 80.0%; Pred. No. 2.3e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgtttgaagcagcag 25
||| ||||| ||||| ||||| |||||
Db 2541 GGAGCTGCTGCTGAGACGACGCTG 2517

PT encoded proteins, useful for treatment of hepatopathies, inflammatory
 PT diseases and proliferative disorders such as cancer -
 XX
 XX
 PS Claim 1; Page 315-316; 392pp; English.

XX The present invention is concerned with the sequence of the genome of the
 CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
 CC the cause of hepatopathies which are not linked to the presence of the
 CC hepatitis A, B and E viruses in man. The genome and proteins of this
 CC virus can be used in gene therapy and vaccination against the virus,
 CC which also causes disorders of the gastrointestinal tract, including
 CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
 CC proliferative disorders such as cancer.
 CC
 SQ Sequence 2040 BP: 692 A; 560 C; 387 G; 401 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2040;
 Best Local Similarity 80.0%; Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggatcgtcgtttgaagcgacagcag 25
 ||| ||||| ||| ||| ||||| |||
 Db 1969 GGAGCTGCTGCTTGATGAGCAGCTG 1945

RESULT 42
 AAN80922
 ID AAN80922 standard; cDNA; 2153 BP.

AC AAN80922;

DT 29-DEC-1990 (first entry)

DE Sequence encoding human estrogen-related receptor protein (hERR2).

KM Hormone receptor; hormone-binding; transcription activation; ss.

OS Homo sapiens.

FI Key Location/Qualifiers
 FT CDS 100..1401
 FT /*tag= a

PN WO8803168-A.

PD 05-MAY-1988.

PF 23-OCT-1987; 87MO-US02782.

PR 20-OCT-1987; 87US-0108471.

PR 04-OCT-1986; 86US-0922585.

PA (SALK) SALK INST FOR BIOL STUD.

PI Evans RM, Weinberger CA, Hollenberg SM, Giguere V;

DR WPI: 1988-133242/19.

DR P-PSDB; AAP80931.

PT Recombinant DNA encoding hormone receptors -
 PT complementing glucocorticoid, mineralocorticoid, thyroid hormone
 PT and novel hormone receptors

PS Claim 16; Fig V-2(B)-1 and -2; 243pp; English.

CC DNAs encoding hormone receptors and the hormone receptors themselves are
 CC claimed. The DNA can be used to make the hormone receptor proteins and
 CC functional modified forms in quantities not previously possible. The
 CC receptor proteins can be used to screen cpds. for receptor-agonist or
 CC receptor-antagonist activity. They can also be used in diagnostic assays.
 CC Also claimed is pure DNA encoding oestrogen-related receptors hERR1 and
 CC hERR2 and hERR1 and hERR2. The new hERR1 and hERR2 receptors will provide

CC the basis for development of an assay system that will lead to the
 CC identification of novel hormones.

SQ Sequence 2153 BP: 445 A; 627 C; 624 G; 457 T; 0 other;

Query Match 68.0%; Score 17; DB 9; Length 2153;
 Best Local Similarity 80.0%; Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggatcgtcgtttgaagcgacagcag 25
 ||| ||||| ||| ||| ||||| |||
 Db 1200 ggacctgctgcatgagcgctgcag 1224

RESULT 43
 AAX06819
 ID AAX06819 standard; DNA; 2526 BP.

AC AAX06819;

DT 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp7 DNA.

KM Omp7; outer membrane protein 7; surface exposed protein; antigen;
 KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

OS Chlamydia pneumoniae.

PN WO9558953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.
 (CHRI/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Mygind P;

DR WPI: 1999-105610/09.

DR P-PSDB; AAW88420.

PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins

PS Claim 6; Page 49-50; 115pp; English.

CC This DNA sequence codes for the novel 89.7 kDa surface exposed
 CC protein Omp7 (see AAW88420) of the human respiratory pathogen
 CC Chlamydia pneumoniae. By generating antibodies against C.
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
 CC was obtained which reacted with outer membrane proteins. The
 CC antibody was used to identify the genes (see AAX06816-27) encoding
 CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of
 CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
 CC the other, and encode polypeptides of about 89.5-100.3 kDa and 15 in
 CC about 56.1 kDa. The invention provides a new species-specific test
 CC for identifying mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The

CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypochloridism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SQ Sequence 312 BP; 88 A; 82 C; 86 G; 55 T; 1 other;

Query Match 68.0%; Score 17; DB 21; Length 312;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcagcag 25
 ||||| || || |||||
 Db 256 GGAATCTCTTGGCTGATCCGACGACG 232

RESULT 38
 AAC07943
 ID AAC07943 standard; cDNA: 394 BP.
 AC AAC07943;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 12018.
 XX
 KW Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 12018: 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC cDNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA⁺ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 394 BP; 71 A; 118 C; 115 G; 83 T; 7 other;

Query Match 68.0%; Score 17; DB 21; Length 394;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcagcag 25
 ||||| || || |||||
 Db 29 ggacatcgtctgttgaagcagcag 53

RESULT 39
 AAC93432
 ID AAC93432 standard; cDNA: 1579 BP.
 AC AAC93432;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein gene 11 SEQ ID NO:21.

XX Human: secreted protein; immunosuppressive; antiarthritic; antineoplastic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW nocotropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; vulnary; autoimmune disease; Rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200061625-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US08981.
 XX
 PR 09-APR-1999; 99US-0128701.
 XX
 PR 20-JAN-2000; 2000US-0177166.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI: 2000-619226/59.
 XX
 DR P-PsDB; AAB51735.
 XX

PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1: Page 417; 500pp; English.

CC Polynucleotide sequences AAC93422 - AAC93449 represent cDNA encoding
 CC human secreted proteins AAB51724 - AAB51777. Sequences AAB51778 -
 CC AAB51825 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences to which they are homologous. The genes and proteins
 CC have activities dependent on the tissues and cells in which they are
 CC expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antineoplastic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nocotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnary. The secreted
 CC proteins, polynucleotides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158332.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159884.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 68.8%; Score 17.2; DB 21; Length 2126;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atctgctgttgaagcagca 24
 DB 320 ATCCGCTTGTGATCGCAGCA 299
 ||| ||||| || ||||| |||

RESULT 34
 AAX25696/c
 ID AAX256969 standard; cDNA; 2203 BP.
 AC AAX256969;
 XX
 DT 12-MAY-2000 (first entry)
 DE Arabidopsis ATP citrate lyase (ACL) B-2 subunit encoding cDNA.
 XX
 KW Arabidopsis; plant plastid; acetyl CoA synthetase; E3 subunit; ACS;
 KW plant plastidic pyruvate dehydrogenase; PPDH; ATP citrate lyase; ACL;
 KW pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;
 KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;
 KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase; ss.
 XX
 OS Arabidopsis sp.
 XX
 PN W0200000619-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-US14382.
 XX
 PI

XX
 PR 26-JUN-1998; 98US-0090717.
 XX
 PA (IOWA) UNIV IOWA STATE RES FOUND INC.
 XX
 PI Nikola BJ, Wurtele ES, Oliver DJ, Behal R, Schnable PS, Ke J;
 PI Johnson JL, Allred CC, Fatland B, Lutziger I, Wen T;
 XX
 DR WPI: 2000-160678/14.
 DR P-PSDB: AAY67408.
 XX
 PT Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase
 PT (PPDH), ATP citrate lyase (ACL), pyruvate decarboxylase (PDC) and
 PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA
 PT levels in plants
 XX
 PS Claim 24; Fig 6; 79pp; English.
 XX
 CC The invention provides nucleic acids encoding Arabidopsis plastidic
 CC acetyl CoA synthetase (ACS), various subunits (specifically the E3
 CC subunit) of plant plastidic pyruvate dehydrogenase (PPDH), the E3
 CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
 CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
 CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
 CC standard recombinant methodology. The ACS, PPDH, ACL, PDC and ALDH
 CC polypeptides, methods and nucleic acid molecules of the invention are
 CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
 CC or organs. A decrease in acetyl CoA is expected to affect the
 CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
 CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
 CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,
 CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
 CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
 CC of acetyl CoA.
 XX
 SQ Sequence 2203 BP; 564 A; 480 C; 494 G; 665 T; 0 other;

Query Match Best Local Similarity 68.8%; Score 17.2; DB 21; Length 2203;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atctgctgttgaagcagca 24
 DB 332 ATCCGCTTGTGATCGCAGCA 311
 ||| ||||| || ||||| |||

RESULT 35
 AAX20559/c
 ID AAX20559 standard; DNA; 5504 BP.
 AC AAX20559;
 XX
 DT 05-MAY-1999 (first entry)
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN W09859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;

PN EPI033405-A2.
XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139466.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154039.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SQ Sequence 466 BP; 98 A; 109 C; 110 G; 149 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 466;
Best Local Similarity 86.4%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Y 2 gatctgctgttgaagcagcagc 23
Db 95 gatctgctgttgaagcagcagc 116

RESULT 31
AAI36883
ID AAI36883 standard; DNA: 466 BP.
XX
AC AAI36883;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #5569 used to measure gene expression in human placenta sample.
XX
KM Probe; microarray; human; placenta; antenatal diagnosis;
XX
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN MO200157272-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001MO-US00663.
XX
PF 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 5569; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
XX
CC The present sequence is one such probe. The probes are useful for
XX
CC producing a microarray for predicting, measuring and displaying gene
XX
CC expression in samples derived from human placenta. The probes are useful
XX
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 466 BP; 98 A; 109 C; 110 G; 149 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 466;
Best Local Similarity 86.4%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Y 2 gatctgctgttgaagcagcagc 23
Db 95 gatctgctgttgaagcagcagc 116

RESULT 32
AAI16338
ID AAI16338 standard; cDNA: 1970 BP.
XX
AC AAI16338;
XX
DT 26-APR-1999 (first entry)
XX
DE Mouse OPF #16 encoding cDNA.
XX
KM Mouse; OPF #16; differentiation inducing activity; bone marrow cell;
XX
KW hydroxyapatite decomposition; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
XX
FT CDS 270..953
XX
FT /tag= a
XX
FT /product= "OPF #16"

JPI1028091-A.

02-FEB-1999.

10-JUL-1997; 973P-0202201.

10-JUL-1997; 973P-0202201.

(SUMU) SUMITOMO SEIYAKU KK.

WPI: 1999-174353/15.

P-PSDB; AAW94508.

A new protein OPF u 16 and gene - useful for differentiation of bone
marrow cells

Claim 2; Page 11-12; 13pp; Japanese.

The present sequence encodes an OPF protein isolated from mouse,
designated OPF #16. OPF #16 has differentiation inducing activity from
a bone marrow cell to a cell having hydroxyapatite decomposing activity.
OPF genes and proteins are useful for screening for inhibitors against
differentiation from bone marrow cells.

Sequence 1970 BP; 471 A; 507 C; 623 G; 369 T; 0 other;

Query Match 68.8%; Score 17.2; DB 20; Length 1970;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Y 4 tctgctgttgaagcagcagc 25
Db 1806 tctgctgttgaagcagcagc 1827

RESULT 33
AAC44857/c
ID AAC44857 standard; DNA: 2126 BP.
XX
AC AAC44857;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 44393.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
XX
KW protein identification; signal transduction pathway;
XX
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.

CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15895 BP; 4641 A; 3798 C; 3736 G; 3720 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15895;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 gatcgtgttgaagcgacag 25
 |||||
 Db 12941 gatcgtgttgaagcgacag 12964

RESULT 29
 AAAS9384
 ID AAAS9384 standard; DNA: 801 BP.
 XX
 AC AAAS9384;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE DNA clone encoding a murine guanine nucleotide releasing factor 4.
 XX
 KW guanine nucleotide releasing factor 4; GRF-4; Ras activator;
 KW Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;
 KW cell differentiation; cell transformation; neuronal disorder; ss.
 OS
 XX
 XX Mus sp.

Location/Qualifiers
 FT 1.801
 FT /*tag= a
 FT /note= "contains 5 internal stop codons and encodes
 FT AAB07793; no termination codon given"

CDS 2..799
 FT /*tag= b
 FT /note= "contains 8 internal stop codons and encodes
 FT AAB07794; no termination codon given"
 FT CDS 3..800
 FT /*tag= c
 FT /note= "contains 5 internal stop codons and encodes
 FT AAB07795; no termination codon given"

PN W0200043510-A2.

XX 27-JUL-2000.

XX 20-JAN-2000; 2000WO-CA00042.

XX 20-JAN-1999; 99CA-2259830.

XX (HSCR-) HSC RES & DEV LP.

XX Rotin D, Pham N;

XX WPI: 2000-499228/44.

XX P-PSDB: AAB07793, AAB07794, AAB07795.

XX Nucleic acids encoding guanine nucleotide releasing factor-4 useful for

XX the treatment of cancers and neuronal disorders -

XX Disclosure; Fig 19B; 89pp; English.

XX The present sequence encodes a murine guanine nucleotide releasing
 CC factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both
 CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate
 CC (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP)
 CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
 CC coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates
 CC Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4

CC is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
 CC the Ras signalling pathway controls numerous cellular functions, such as
 CC cell metabolism, proliferation, differentiation and transformation.
 CC Therefore modulation of Ras activity may provide a mechanism for
 CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
 CC in the treatment of diseases associated with inappropriate GRF4
 CC expression and activity such as cancers and neuronal disorders. The
 CC GRF4 polypeptides may be used as antigens in the production of
 CC antibodies against GRF4 and in assays to identify modulators (agonists
 CC and antagonists) of GRF4 expression and activity. The anti-GRF4
 CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
 CC expression and activity. Inhibition of Ras can reduce cellulose
 CC proliferation and cancers.
 CC
 SQ Sequence 801 BP; 173 A; 266 C; 232 G; 130 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 801;
 Best Local Similarity 94.7%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 tctgctgttgaagcgacag 22
 |||||
 Db 525 tctgctgttgaagcgacag 543

RESULT 30
 AA115487
 ID AA115487 standard; DNA: 466 BP.
 XX
 AC AA115487;

XX 12-OCT-2001 (first entry)

XX Probe #5420 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX Homo sapiens.

XX W0200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID NO 5420; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.

XX DR WPI: 1999-580441/49.
 XX PT New vaccines for measles and respiratory syncytial virus (RSV)
 XX PS Disclosure: Page 120-126; 171pp; English.
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Ser134Tyr; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 SQ Sequence 15894 BP: 4638 A; 3798 C; 3735 G; 3723 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgttgaagcgcagcag 25
 Db 12940 gatctgctgttagaatagcacag 12963

RESULT 27
 AA222908
 ID AA222908 standard; DNA: 15894 BP.
 XX AC AA222908;
 XX 10-JAN-2000 (first entry)
 XX DE Nucleotide sequence of AIR-C vaccine measles virus genome.
 XX ME Measles virus: attenuated; human respiratory syncytial virus; RSV;
 XX KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX OS Synthetic.
 XX OS Measles virus.
 XX PN WO9949017-A2.
 XX PD 30-SEP-1999.
 XX PF 22-MAR-1999: 99WO-US06225.
 XX PR 26-MAR-1998: 98US-0079466.
 XX PA (AMCY) AMERICAN CYANAMID CO.
 XX PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX DR WPI: 1999-580441/49.
 XX PT New vaccines for measles and respiratory syncytial virus (RSV)
 XX PS Disclosure: Page 126-131; 171pp; English.
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Ser134Tyr; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.

CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal,
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 SQ Sequence 15894 BP: 4644 A; 3793 C; 3731 G; 3726 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgttgaagcgcagcag 25
 Db 12940 gatctgctgttagaatagcacag 12963

RESULT 28
 AA222906
 ID AA222906 standard; DNA: 15895 BP.
 XX AC AA222906;
 XX 10-JAN-2000 (first entry)
 XX DE Nucleotide sequence of Moraten vaccine measles virus genome.
 XX ME Measles virus: attenuated; human respiratory syncytial virus; RSV;
 XX KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX OS Synthetic.
 XX OS Measles virus.
 XX PN WO9949017-A2.
 XX PD 30-SEP-1999.
 XX PF 22-MAR-1999: 99WO-US06225.
 XX PR 26-MAR-1998: 98US-0079466.
 XX PA (AMCY) AMERICAN CYANAMID CO.
 XX PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX DR WPI: 1999-580441/49.
 XX PT New vaccines for measles and respiratory syncytial virus (RSV)
 XX PS Disclosure: Page 114-120; 171pp; English.
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal,
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated

CC least 1 attenuating mutation in the 3' genomic promoter region and at
 CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
 CC can be used as a vaccine to immunise an individual against such a virus.
 XX
 SQ Sequence 15894 BP; 4639 A; 3791 C; 3729 G; 3735 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatcgcgtgttgaagcgcagcag 25
 |||||
 Db 12940 gatcgcgtgttgaagcgcagcag 12963

RESULT 22
 AA222901
 ID AA222901 standard; DNA: 15894 BP.
 XX
 AC AA222901;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Nucleotide sequence of wild type measles virus Edmonston strain.
 XX
 KW Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KW mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Measles virus.
 XX
 PN WO9949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PT New vaccines for measles and respiratory syncytial virus (RSV)
 PS
 XX Disclosure: Page 86-92; 171pp; English.
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr724Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr419Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15894 BP; 4639 A; 3800 C; 3737 G; 3718 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatcgcgtgttgaagcgcagcag 25
 |||||
 Db 12940 gatcgcgtgttgaagcgcagcag 12963

RESULT 23
 AA222902
 ID AA222902 standard; DNA: 15894 BP.
 XX
 AC AA222902;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Nucleotide sequence of wild type measles virus 1977 strain.
 XX
 KW Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KW mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Measles virus.
 XX
 PN WO9949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PT New vaccines for measles and respiratory syncytial virus (RSV)
 PS
 XX Disclosure: Page 92-97; 171pp; English.
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr724Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr419Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15894 BP; 4669 A; 3780 C; 3716 G; 3729 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatcgcgtgttgaagcgcagcag 25
 |||||
 Db 12940 gatcgcgtgttgaagcgcagcag 12963

RESULT 24
 AA222903
 ID AA222903 standard; DNA: 15894 BP.
 XX
 AC AA222903;
 XX

```

XX (AMCY ) AMERICAN CYANAMID CO.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA:
XX WPI: 1998-230710/20.
XX P-PSDB: AAW48710.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
XX attenuating mutation in 3' genomic promoter region and RNA
XX polymerase gene, useful as vaccine to immunise against such virus
XX
XX Disclosure: Page 217-228; 426pp; English.
XX
XX This sequence encodes the L protein from Measles virus ATK-C vaccine.
CC This sequence is used in a method which involves the isolation of
CC recombinantly-generated, attenuated, non-segmented, negative-sense,
CC single stranded RNA virus of the order Mononegavirales which have at
CC least 1 attenuating mutation in the 3' genomic promoter region and at
CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC can be used as a vaccine to immunise an individual against such a virus.
XX
XX Sequence 15894 BP; 4644 A; 3793 C; 3731 G; 3726 T; 0 other:
SQ

```

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgcaagcag 25
 |||||
 Db 12940 gatctgctgtagaatagcaacag 12963

RESULT 20
 AAV18265 standard; DNA; 15894 BP.
 ID AAV18265;
 AC AAV18265;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus 1977 isolate wild-type L protein genomic DNA.
 XX
 KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 KM single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 XX Key Location/Qualifiers
 FH 9234..15785
 FT CDS /*tag= a
 FT /*product= "L protein".
 XX
 W09813501-A2.
 PN
 PD 02-APR-1998.
 XX
 PF 19-SEP-1997; 97WO-US16718.
 XX
 PR 27-SEP-1996; 96US-0026823.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX WPI: 1998-230710/20.
 DR P-PSDB: AAV18265.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,

```

PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
XX
XX Claim 31; Page 105-116; 426pp; English.
XX
XX This sequence encodes the L protein from Measles virus isolate 1977.
CC This sequence is used in a method which involves the isolation
CC of recombinantly-generated, attenuated, non-segmented, negative-sense,
CC single stranded RNA virus of the order Mononegavirales which have at
CC least 1 attenuating mutation in the 3' genomic promoter region and at
CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC can be used as a vaccine to immunise an individual against such a virus.
XX
XX Sequence 15894 BP; 4669 A; 3780 C; 3716 G; 3729 T; 0 other:
SQ

```

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgcaagcag 25
 |||||
 Db 12940 gatctgctgtagaatagcaacag 12963

RESULT 21
 AAV18266 standard; DNA; 15894 BP.
 ID AAV18266;
 AC AAV18266;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus 1983 isolate wild-type L protein genomic DNA.
 XX
 KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 KM single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 XX Key Location/Qualifiers
 FH 9234..15785
 FT CDS /*tag= a
 FT /*product= "L protein".
 XX
 W09813501-A2.
 PN
 PD 02-APR-1998.
 XX
 PF 19-SEP-1997; 97WO-US16718.
 XX
 PR 27-SEP-1996; 96US-0026823.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX WPI: 1998-230710/20.
 DR P-PSDB: AAW48705.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 XX
 XX Claim 31; Page 124-134; 426pp; English.
 XX
 XX This sequence encodes the wild-type L protein from Measles virus isolate
 CC 1983. This sequence is used in a method which involves the isolation
 CC of recombinantly-generated, attenuated, non-segmented, negative-sense,
 CC single stranded RNA virus of the order Mononegavirales which have at

DE Measles virus Moraten vaccine L protein genomic DNA.
XX L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus; Mononegavirales; ss.
XX
OS Measles virus.
XX
FH Key Location/Qualifiers
FT CDS 9234..15785
FT /tag= a
FT /product= "L protein"
XX
PN WO9813501-A2.
XX
PD 02-APR-1998.
XX
PF 19-SEP-1997; 97WO-US16718.
XX
PR 27-SEP-1996; 96US-0026823.
XX
PA (AMCY) AMERICAN CYANAMID CO.
PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
PI P-PSDB; AAW48708.
DR WPI: 1998-230710/20.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
XX
PS Claim 31: Page 180-190; 426pp; English.
XX
CC This sequence encodes the L protein from Measles virus Moraten
CC vaccine. This sequence is used in a method which involves the
CC isolation of recombinantly-generated, attenuated, non-segmented,
CC negative-sense, single stranded RNA virus of the order Mononegavirales
CC which have at least 1 attenuating mutation in the 3' genomic promoter
CC region and at least 1 attenuating mutation in the RNA polymerase gene.
CC This RNA virus can be used as a vaccine to immunise an individual against
CC such a virus.
XX
SQ Sequence 15894 BP: 4640 A; 3798 C; 3736 G; 3720 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
Db 12940 gatctgctgttgaagcgacag 12963

RESULT 18
AAV18270
ID AAV18270 standard; DNA: 15894 BP.
XX
AC AAV18270;
XX
DT 13-OCT-1998 (first entry)
XX
DE Measles virus Zagreb vaccine L protein genomic DNA.
XX
KM L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus; Mononegavirales; ss.
XX
OS Measles virus.
XX
FH Key Location/Qualifiers
FT CDS 9234..15785

FT /tag= a
FT /product= "L protein"
XX
PN WO9813501-A2.
XX
PD 02-APR-1998.
XX
PF 19-SEP-1997; 97WO-US16718.
XX
PR 27-SEP-1996; 96US-0026823.
XX
PA (AMCY) AMERICAN CYANAMID CO.
PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
PI P-PSDB; AAW48709.
DR WPI: 1998-230710/20.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
XX
PS Claim 31: Page 199-209; 426pp; English.
XX
CC This sequence encodes the L protein from Measles virus Zagreb vaccine.
CC This sequence is used in a method which involves the isolation of
CC recombinantly-generated, attenuated, non-segmented, negative-sense,
CC single stranded RNA virus of the order Mononegavirales which have at
CC least 1 attenuating mutation in the 3' genomic promoter region and at
CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC can be used as a vaccine to immunise an individual against such a virus.
XX
SQ Sequence 15894 BP: 4638 A; 3798 C; 3735 G; 3723 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
Db 12940 gatctgctgttgaagcgacag 12963

RESULT 19
AAV18271
ID AAV18271 standard; DNA: 15894 BP.
XX
AC AAV18271;
XX
DT 13-OCT-1998 (first entry)
XX
DE Measles virus AIK-C vaccine L protein genomic DNA.
XX
KM L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus; Mononegavirales; ss.
XX
OS Measles virus.
XX
FH Key Location/Qualifiers
FT CDS 9234..15785
FT /tag= a
FT /product= "L protein"
XX
PN WO9813501-A2.
XX
PD 02-APR-1998.
XX
PF 19-SEP-1997; 97WO-US16718.
XX
PR 27-SEP-1996; 96US-0026823.

CC eukaryotic cells to a cytokine by regulating the tyrosine kinase activity
CC of Jak kinases. Jak kinases mediate cytokine activity through their
CC tyrosine phosphorylation in response to cytokine-receptor binding. This
CC method is useful in treating diseases caused by an excessive cellular
CC response to a Jak kinase mediated cytokine, by inhibiting Jak kinase
CC activity. Such diseases include those caused by excessive proliferation
CC of cells. Treatment of disease conditions caused by a deficient cellular
CC response or non-responsiveness to a cytokine involves enhancing Jak
CC kinase activity. Jak kinase sequence is useful in identifying its
CC inhibitors.
XX
SQ Sequence 3561 BP: 699 A; 1113 C; 1097 G; 652 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 3561;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 gatctgctgttgaagcgcagcag 25
|||||
Db 3021 GATCTGCTGGCGCAAGACGACGAG 2998

RESULT 13
AAH98351/C
ID AAH98351 standard; cDNA: 4218 BP.
XX
AC AAH98351;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SFO ID NO: 208.
XX
KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
XX
DR P-PSDB; AAM23692.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 322-323; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX

SQ Sequence 4218 BP: 843 A; 1300 C; 1283 G; 792 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 4218;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 gatctgctgttgaagcgcagcag 25
|||||
Db 3364 GATCTGCTGGCGCAAGACGACGAG 3341

RESULT 14
AAQ40480
ID AAQ40480 standard; cDNA: 15894 BP.
XX
AC AAQ40480;
XX
DT 19-AUG-1993 (first entry)
XX
DE Attenuated measles virus strain AIK-C.
XX
KW Paramyxoviridae; RNA virus; attenuation; ss.
XX
OS Human measles virus.
XX
FH Location/Qualifiers
XX
FT CDS 108..1685
FT /*tag= a
FT /*label= N
FT 1807..3330
FT /*tag= b
FT /*label= P/C
FT 3438..4445
FT /*tag= c
FT /*label= M
FT 5458..7110
FT /*tag= d
FT /*label= F
FT 7271..9124
FT /*tag= e
FT /*label= H
FT 9233..15785
FT /*tag= f
FT /*label= L
XX
PN EP540135-A.
XX
PD 05-MAY-1993.
XX
PE 10-MAR-1992; 92EP-0302004.
XX
PR 14-OCT-1991; 91JP-0293625.
XX
PA (KITA) KITASATO INST.
PA (KITA) KITASATO KENKYUSHO SH.
XX
PI Makino S, Mori T, Sasaki K;
XX
DR WPI: 1993-145503/18
DR P-PSDB; AAR34537, AAR34538, AAR34539, AAR34540, AAR34541, AAR39592.
XX
PT New attenuated measles vaccine virus strain - retains high
PT immunogenicity with reduced pyrogenicity and having no
PT neurological complications
XX
PS Claim 1; Page 4-37; 47pp; English.
XX
CC This sequence represents the entire genome of the attenuated measles
CC virus strain AIK-C. The virus retains high immunogenicity with
CC reduced pyrogenicity and no neurological complications.
XX
SQ Sequence 15894 BP: 4649 A; 3795 C; 3726 G; 3724 T; 0 other;

XX Inhibiting a cellular response to a cytokine by inhibiting Jak
 PT kinase - to treat diseases caused by excessive response to
 PT cytokine, e.g. erythrocytosis and other cellular proliferative
 PT diseases
 XX Claim 29; Fig 3A,C,B,D,E; 167pp; English.
 PS
 CC Inhibiting the activity of a jak kinase (pref. Jak1, jak2, jak3 or
 CC Tyk2) in a eukaryotic cell is claimed as a method of inhibiting the
 CC biological response of that cell to a cytokine (not IL-3 or
 CC erythropoietin). The present sequence (human TYK2 kinase) was
 CC published by Fimbach-Kraft et al., Oncogene 5:1329-1336 (1990).
 CC It encodes a 1187 amino acid protein which includes the epitopic
 CC sequence SPSEKHFYQGRHRPEPS (amino acids 819-837). Antibodies which
 CC selectively bind this epitope are able to bind Tyk2 without
 CC interfering with the activity of the kinase. Such antibodies are
 CC claimed and are useful for detecting and extracting Tyk2.
 XX
 SQ Sequence 3561 BP; 699 A; 1112 C; 1098 G; 652 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 16; Length 3561;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgtttgaagcgagcag 25
 DB 3021 GATCTGCTGGCGAAGAGCAGCAG 2998
 |||||
 RESULT 11
 AAC66246/c
 ID AAC66246 standard; DNA: 3561 BP.
 AC AAC66246;
 XX
 DT 19-FEB-2001 (first entry)
 XX
 DE Human tyk2 kinase nucleotide sequence.
 XX
 KW Jak3; kinase; cytokine; cellular response; inhibition; tyk2; ds;
 KW cell proliferation; erythrocytosis.
 OS Homo sapiens.
 XX
 PN US6136595-A.
 PD 24-OCT-2000.
 XX
 PF 18-JUN-1996; 96US-0665574.
 XX
 PR 29-JUL-1994; 94US-0282012.
 PR 29-JUL-1993; 93US-0097997.
 PR 09-SEP-1993; 93US-0118968.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 PI Silvennoinen O, Wiltuhn BA, Ihle J;
 DR WPI: 2000-686080/67.
 DR P-PSDB; AAB35726.
 XX
 PT New DNA encoding Jak3 kinase is useful as cytokine regulator for
 PT treating cell proliferation -
 PS Disclosure: Fig 3; 100pp; English.
 XX
 CC This invention relates to DNA encoding a murine Jak3 protein. The amino
 CC acid sequence of the Jak3 protein is given in AAB35715. The Jak family
 CC of kinases are involved in the cellular response to the binding of
 CC cytokines to their respective receptors. Jak3 kinase mediated activation
 CC of some cytokines through their phosphorylation in response to

CC cytokine-receptor binding. Inhibiting the activity of Jak3 kinase (at the
 CC nucleic acid level with antisense sequences or ribozymes, or at the
 CC protein level with antibodies, kinase inhibitors etc.) is used to treat,
 CC or diagnose, diseases caused by excessive secretion of certain cytokines,
 CC e.g. excessive cell proliferation such as erythrocytosis. Alternatively,
 CC the Jak3 protein and polynucleotide can be used to treat conditions
 CC associated with defective Jak3 activity. The DNA sequence can be used to
 CC produce recombinant Jak3 and this used to raise antibodies useful as
 CC specific inhibitors or to detect or isolate Jak3 without interfering
 CC with its enzymatic activity. The present sequence represents DNA encoding
 CC a tyk2 protein used in the isolation and characterisation of the jak3
 CC protein of the invention.
 XX
 SQ Sequence 3561 BP; 699 A; 1113 C; 1097 G; 652 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 21; Length 3561;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgtttgaagcgagcag 25
 DB 3021 GATCTGCTGGCGAAGAGCAGCAG 2998
 |||||
 RESULT 12
 AAD03609/c
 ID AAD03609 standard; cDNA: 3561 BP.
 AC AAD03609;
 XX
 DT 19-JUN-2001 (first entry)
 XX
 DE Human Tyk2 kinase cDNA.
 XX
 KW Human; Janus kinase; Jak; therapy; cytokine; cellular response;
 KW antiproliferative; cytostatic; cell proliferative disorder;
 KW tyrosine kinase; Tyk2 kinase; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..3561
 FT CDS
 FT /*tag= a
 FT /*product= "Human Tyk2 kinase"
 FT /*note= "CDS does not include stop codon"
 FT /*partial
 XX
 PN US6210654-B1.
 PD 03-APR-2001.
 XX
 PF 08-OCT-1997; 97US-0946994.
 XX
 PR 18-JUN-1996; 96US-0665574.
 PR 29-JUL-1993; 93US-0097997.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S HOSPITAL.
 XX
 PI Ihle J, Wiltuhn BA, Quelle FW, Silvennoinen O;
 DR WPI: 2001-265367/27.
 DR P-PSDB; AAE00354.
 XX
 PT Modulating a biological response mediated by Jak kinase 2 activation to
 PT a cytokine, useful for treating excessive proliferation of eukaryotic
 PT cells, comprises inhibiting or enhancing tyrosine kinase activity of
 PT Jak kinase in the cell -
 PS Example 1; Fig 3; 100pp; English.
 XX
 CC The present sequence is human tyrosine kinase (Tyk2) cDNA.
 CC The present invention relates to a method for modulating response of

DE Human Lhx3 genomic DNA SEQ ID NO:22.
 XX
 XX Lhx3; LIM-3; P-LIM; identification; characterisation; diagnosis;
 KW chromosome 9; pituitary disease; subtelomeric region; mutation;
 KW pituitary trophic hormone gene promoter; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200050868-A2.
 XX
 PD 31-AUG-2000.
 XX
 PF 22-FEB-2000; 2000WO-US04424.
 XX
 PR 22-FEB-1999; 99US-0121110.
 XX
 PA (ADRE-) ADVANCED RES & TECHNOLOGY INST.
 XX
 PI Rhodes SJ, Bridwell JL, Meier BC, Parker GE, Price JR;
 PI Showalter AD, Sloop KW;
 XX
 DR WPI; 2000-594085/56.
 XX
 PT New isolated nucleic acid encoding mammalian Lhx3 for identifying a
 PT human with a disease, disorder, or condition caused by an altered level
 PT of expression or binding of Lhx3 -
 XX
 PS Claim 30; Fig 26; 239pp; English.
 XX
 CC The present invention describes an isolated nucleic acid (1) encoding a
 CC mammalian Lhx3. (1) is used in assays to: (1) detect and quantify the
 CC presence and level of expression of Lhx3, Lhx3a or Lhx3b, in a sample;
 CC (2) identify a compound that affects expression, the level of expression,
 CC or the activity of Lhx3, Lhx3a, or Lhx3b in a cell; (3) identify a
 CC compound that affects binding of Lhx3 to nucleic acid or Lhx3 induction
 CC of a pituitary trophic hormone gene promoter; (4) identify a human
 CC afflicted with a disease, disorder, or condition caused by altered
 CC expression of Lhx3 or altered level of binding of Lhx3 to a nucleic acid;
 CC and (5) detect a mutation in a Lhx3 allele in a human. The coding region
 CC of human Lhx3 has been genomically mapped to the subtelomeric region of
 CC chromosome 9. Lhx3 is also known as P-LIM or LIM-3. The present sequence
 CC represents the specifically claimed human Lhx3 genomic DNA sequence
 CC given in the present invention.
 XX
 SQ Sequence 8867 BP; 1462 A; 2874 C; 2940 G; 1589 T; 2 other;

Query Match 74.4%; Score 18.6; DB 21; Length 8867;
 Best Local Similarity 84.0%; Pred. No. 53;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttaagcgcagcag 25
 ||||||||| ||||||| ||||
 Db 4007 GGATCTGCTCTGTGAAGCCGCGCAG 3983

RESULT 9
 AAX20711
 ID AAX20711 standard; DNA; 808 BP.
 XX
 AC AAX20711;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN WO9859034-A2.
 XX
 DR

PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 XX
 DR WPI; 1999-081273/07.
 XX
 PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX
 PS Claim 1; Page 864; 1150pp; English.
 XX
 CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX
 SQ Sequence 808 BP; 204 A; 200 C; 221 G; 180 T; 3 other;

Query Match 70.4%; Score 17.6; DB 20; Length 808;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gattcgtctgttaagcgcagcag 25
 ||| ||||| ||| ||||||| |||
 Db 499 gatgctcttttaagcgcagcag 522

RESULT 10
 AAO85414/C
 ID AAO85414 standard; DNA; 3561 BP.
 XX
 AC AAO85414;
 XX
 DT 06-OCT-1995 (first entry)
 XX
 DE Human TYK2 kinase coding sequence.
 XX
 KW JAK family; protein tyrosine kinase; cytokine receptor;
 KW phosphorylation; signal transduction; activation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..3561
 FT /*tag= a
 FT /*product= human_TYK2
 FT /*note= "nucleotides 817..1680 shown as Fig. 3C
 and nucleotides 1681..2544 shown as Fig. 3B"

WO9503701-A.
 XX
 PD 09-FEB-1995.
 XX
 PF 29-JUL-1994; 94WO-US08676.
 XX
 PR 29-JUL-1993; 93US-0097997.
 XX
 PA (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 PI Ihle JN, Quelle FW, Silvenoinen O, Witthuhn BA;
 XX
 DR WPI; 1995-081950/11.
 DR P-PSDB; AAR70832.

XX 24-JUL-2001 (first entry)
 XX Legionella pneumophila nucleotide sequence SEQ ID NO:322.
 DE
 XX
 XX Species specific; genus specific; family specific; probe; detection;
 KW identification; algal; archaeal; bacterial; fungal; parasitica;
 KW microorganism; diagnosis; translation elongation factor Tu; toxin;
 KW translation elongation factor G; RecA recombinase; resistance;
 KW catalytic subunit of proton-translocating ATPase; antimicrobial;
 KW vaccine; primer; ds.
 XX
 XX Legionella pneumophila.
 OS
 PN WO200123604-A2.
 PD 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-CA01150.
 PF
 XX 28-SEP-1999; 99CA-2283458.
 PR 19-MAY-2000; 2000CA-2307010.
 XX
 PA (INFE-) INFECTIO DIAGNOSTIC (TDI) INC.
 XX
 PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
 PI Picard FT, Roy PH;
 PI
 XX WPI; 2001-245006/25.
 DR
 PT Nucleic acid sequences are used to generate universal probes and
 PT primers which can be used to identify and detect the presence of algal,
 PT archaeal, bacterial, fungal and parasitica species in a test sample -
 XX
 XX Claim 25; Page 597; 1580pp; English.
 PS
 CC The present invention describes a method for generating a repertory of
 CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes
 CC and/or primers are derived. The method comprises amplifying the nucleic
 CC acids of determined algal, archaeal, bacterial, fungal and parasitica
 CC species with a combination of defined primer pairs. The method can be
 CC used for producing probes and/or primers for detecting one or more
 CC related microorganisms e.g. algae, archaea, bacteria, fungi and
 CC parasites, for universal detection and for specific and ubiquitous
 CC detection and identification of an algal, archaeal, bacterial, fungal
 CC and parasitica species, genus, family and group. A nucleic acid (I)
 CC obtained using the method of the invention can be used for the universal
 CC detection of any bacterium, fungus or parasite in a sample and for the
 CC detection of at least one antimicrobial agent resistance gene or at
 CC least one toxin gene. hexa nucleic acids are used for the specific and
 CC ubiquitous detection and for identification of Streptococcus pneumoniae.
 CC (1) can be used to design a therapeutic agent which is effective against
 CC microorganisms. Microbial species or genus or family or phylum or group
 CC which can be detected include Abiotrophia adiacens, Bordetella sp.,
 CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
 CC Mycobacteriaceae family, Pseudomonas group, Streptococcus sp.,
 CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
 CC provides faster results than substrate specificity tests as results can
 CC be determined in an hour and improved accuracy is also achieved.
 CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
 CC which are given in the exemplification of the present invention.
 CC
 CC Sequence 824 BP; 226 A; 142 C; 221 G; 235 T; 0 other;
 SO

Query Match 76.8%; Score 19.2; DB 22; Length 824;
 Best Local Similarity 87.5%; Pred. No. 20;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctgcgtttgaagcagcagcag 25
 ||||||| ||||||| |||||||
 DB 54 gatctggtttgaagtcagcagcag 77

RESULT 7
 AAA92065/c
 ID AAA92065 standard; cDNA: 2515 BP.
 XX
 XX AAA92065;
 AC
 XX 04-JAN-2001 (first entry)
 DT
 DE Human Lhx3c nucleic acid sequence SEQ ID NO:25.
 XX
 XX Lhx3: LIM-3; P-LIM; identification; characterisation; diagnosis;
 KW chromosome 9; pituitary disease; subtelomeric region; mutation;
 KW pituitary trophic hormone gene promoter; ss.
 XX
 XX Homo sapiens.
 OS
 PN WO200050868-A2.
 PD 31-AUG-2000.
 XX
 XX 22-FEB-2000; 2000WO-US04424.
 PF
 XX 22-FEB-1999; 99US-0121110.
 PR
 XX (ADRE-) ADVANCED RES & TECHNOLOGY INST.
 PA
 XX Rhodes SJ, Bridwell JL, Meier BC, Parker GE, Price JR;
 PI Showalter AD, Sloop KW;
 PI
 XX WPI; 2000-594085/56.
 DR
 PT New isolated nucleic acid encoding mammalian Lhx3 for identifying a
 PT human with a disease, disorder, or condition caused by an altered level
 PT of expression or binding of Lhx3 -
 XX
 XX Example 4; Fig 19; 239pp; English.
 PS
 CC The present invention describes an isolated nucleic acid (I) encoding a
 CC mammalian Lhx3. (1) is used in assays to: (1) detect and quantify the
 CC presence and level of expression of Lhx3, Lhx3a or Lhx3b, in a sample;
 CC (2) identify a compound that affects expression, the level of expression,
 CC or the activity of Lhx3, Lhx3a, or Lhx3b in a cell; (3) identify a
 CC compound that affects binding of Lhx3 to nucleic acid or Lhx3 induction
 CC of a pituitary trophic hormone gene promoter; (4) identify a human
 CC afflicted with a disease, disorder, or condition caused by altered
 CC expression of Lhx3 or altered level of binding of Lhx3 to a nucleic acid;
 CC and (5) detect a mutation in a Lhx3 allele in a human. The coding region
 CC of human Lhx3 has been genomically mapped to the subtelomeric region of
 CC chromosome 9. Lhx3 is also known as P-LIM or LIM-3. The present sequence
 CC represents the human Lhx3c nucleic acid sequence which is used in an
 CC example from the present invention.
 CC
 CC Sequence 2515 BP; 454 A; 819 C; 760 G; 482 T; 0 other;
 SO

Query Match 74.4%; Score 18.6; DB 21; Length 2515;
 Best Local Similarity 84.0%; Pred. No. 44;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatctgcgtttgaagcagcagcag 25
 ||||||| ||||||| |||||||
 DB 183 GGATCTGCTCTGTGAAGCCGCGCAG 159

RESULT 8
 AAA92058/c
 ID AAA92058 standard; DNA: 8867 BP.
 XX
 XX AAA92058;
 AC
 XX 04-JAN-2001 (first entry)
 DT
 XX


```

AC AA062386;
XX
XX 16-NOV-1994 (first entry)
DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
XX Biotin: expression: enterobacteria; vitamin H; synthesis:
KM plasmid: pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KM promoter plac; biotin synthase; KAPA synthase;
KM 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KM delthiobiotin synthase; DAPA synthase;
KM S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW seborrhoea; dermatitis; ds.
XX
XX Escherichia coli DSM498.
XX
XX Key Location/Qualifiers
FH 1..96
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT 23..28
FT /tag= b
FT /standard_name= "promoter plac"
FT 45..50
FT /tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT 105..109
FT /tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "bioB RBS no. g"
FT 117..1157
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /tag= h
FT /standard_name= "bioC RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioC"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "bioD RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "DTB synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioD15"
FT /number= 4
FT /standard_name= "dethiobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "bioD15 substitution"
FT 3742..3746
RBS

```

```

FT /tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine; 8-amino-
FT 5088..5093
FT /tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT CDS
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT terminator
FT /tag= q
FT /standard_name= "rho-independent transcriptional
FT 5583..5605
FT /tag= r
FT stem_loop
FT
FT WO9408023-A.
FT
FT 14-APR-1994.
FT
FT 01-OCT-1993; 93MO-EP02688.
FT
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT
FT (LON2 ) LONZA AG.
FT
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT
FT WPI: 1994-135587/16.
FT DR P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT
FT Biochemical biotin prodn. using enterobacterial biotin-gene
FT PT providing vitamin H in high yield
FT
FT Claim 1: Fig 6, Page 47-55 and 60-65; 92pp; German.
FT
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT CC of biotin, arranged in a transcription unit. Microorganisms
FT CC contg. these DNA fragments or plasmids may be used in the prodn.
FT CC of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT CC loss of appetite and tiredness.
FT
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT Query Match 100.0%; Score 25; DB 15; Length 5872;
FT Best Local Similarity 100.0%; Pred. NO. 0.068;
FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT Oy 1 ggaatcgtcgttgaagcagcag 25
FT |
FT Db 179 ggaatcgtcgttgaagcagcag 203
FT
FT RESULT 6
FT AAH00331
FT ID AAH00331 standard; DNA; 824 BP.
FT XX
FT AC AAH00331;

```

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA;
 XX
 XX WPI; 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 XX Transgenic plants with high biotin levels - transformed with DNA
 PT encoding diaminopelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2; Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 XX
 XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.052;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcagcag 25
 ||||||||||||||||||||
 Db 63 ggaatcgtctgttgaagcagcag 87

RESULT 3

AAAN91329 standard; DNA: 1084 BP.

XX
 AC AAN91329;

XX 15-FEB-1990 (first entry)

XX E.coli Bio B gene.

XX E.coli: Bio B gene; biotin.

XX Escherichia coli.

XX Key Location/Qualifiers
 FH 24..1064
 FT CDS /*tag=a

XX GB2216530-A.

XX 11-OCT-1989.

XX 17-MAR-1989; 89GB-0006210.

XX 22-MAR-1988; 88GB-0006804.

XX 17-MAR-1989; 89GB-0006210.

XX (UKAG-) UK MIN. AGRIC. FISH.

XX Pearson BM, McKee RA;

XX WPI; 1989-295085/41. P-PSDB P91392

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes

PT - derived from E.coli and capable of replication and expression in other

PT microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

XX The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 XX
 XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcagcag 25
 ||||||||||||||||||||
 Db 86 ggaatcgtctgttgaagcagcag 110

RESULT 4

AAAN60496 standard; DNA: 1121 BP.

XX
 AC AAN60496;

XX 17-OCT-1991 (first entry)

XX Sequence encoding biotin synthesising enzyme.

XX Biotin synthetic enzyme; E.coli; deshtobiotin; ds.

XX Key Location/Qualifiers
 FH 42..1082
 FT CDS /*tag=a

XX JP61149091-A.

XX 07-JUL-1986.

XX 24-DEC-1984; 84JP-0272605.

XX 24-DEC-1984; 84JP-0272605.

XX (NIPS) NIPPON SODA KK.

XX WPI; 1986-216622/33.

XX P-PSDB; AAP60536.

XX Double stranded DNA encoding biotin synthesising enzyme -

PT comprises transformed mutant E.coli strain contg. cyclic doubled

PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

XX Disclosure: Page 534; 23pp; Japanese.

XX The sequence may be expressed by a transformed E.coli host, cultured

CC in a medium containing deshtobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcagcag 25
 ||||||||||||||||||||
 Db 104 ggaatcgtctgttgaagcagcag 128

RESULT 5

AAO62386 standard; DNA: 5872 BP.

XX

85 16.2 64.8 852 22 AAE67926
 86 16.2 64.8 1088 9 AAN80167
 87 16.2 64.8 1280 21 AAC77626
 88 16.2 64.8 1338 21 AAC93411
 89 16.2 64.8 1388 21 AAA27388
 90 16.2 64.8 1478 22 AAS01501
 91 16.2 64.8 1657 19 AAV44828
 92 16.2 64.8 1694 17 AAV15232
 93 16.2 64.8 1694 18 AAV94635
 94 16.2 64.8 1725 19 AAV9336
 95 16.2 64.8 1725 21 AAA54599
 96 16.2 64.8 1743 20 AA233514
 97 16.2 64.8 1743 20 AA233514
 98 16.2 64.8 1804 22 AAS06020
 99 16.2 64.8 1810 18 AAT71704
 100 16.2 64.8 1910 19 AAV69335

ALIGNMENTS

RESULT 1

AA162941
 ID AA162941 standard; DNA: 839 BP.

AC AA162941:

DT 22-OCT-2001 (first entry)

DE Human genomic DNA SEQ ID NO 269.

XX Human; noctropic; neuroprotective; cytostatic; dermatological; virucide;
 KW antiparkinsonian; antisticking; antianaemic; antithrombotic; cancer;
 KW antineumatic; hepatotropic; cerebroprotective; antineuroinflammatory;
 KW antiparasitic; cardiac; antidiabetic; anticonvulsant; antifungal;
 KW neurological disease; infection; nephrotoxic; gene therapy; vaccine;
 ds.

OS Homo sapiens.

PN WO200155449-A1.

PD 02-AUG-2001.

PE 17-JAN-2001: 2001WO-US01346.

XX 31-JAN-2000: 2000US-0179065.
 PR 04-FEB-2000: 2000US-0180628.
 PR 19-MAY-2000: 2000US-0205115.
 PR 07-JUL-2000: 2000US-0216880.
 PR 14-JUL-2000: 2000US-0218290.
 PR 14-AUG-2000: 2000US-0225447.
 PR 01-SEP-2000: 2000US-0230437.
 PR 06-SEP-2000: 2000US-0234997.
 PR 25-SEP-2000: 2000US-0236367.
 PR 13-OCT-2000: 2000US-0239937.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 17-NOV-2000: 2000US-0246528.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249214.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250191.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.

PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251989.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM.
 DR WPI: 2001-476225/51.
 XX
 PT Novel plasma membrane associated proteins useful for diagnosing,
 PT treating, preventing and/or prognosing disorders related to the
 PT proteins, including cancer, immune response and neuronal disorders -
 XX
 PS Example 2: SEQ ID NO 269; 532pp + Sequence Listing; English.

XX The invention relates to novel genes (AA162752-AA162961) and proteins
 CC (AA162347-AA162415) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 839 BP: 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
 Best Local Similarity 100.0%; Pred. No. 0.051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ggaatcgtgttgaagcagcagcag 25
 ||||||||||||||||||||
 Db 95 ggaatcgtgttgaagcagcagcag 119

RESULT 2

AA162941
 ID AA162941 standard; DNA: 1041 BP.

AC AA162941:

DT 12-APR-1999 (first entry)

DE E. coli biotin synthetase (BioB) coding sequence.

KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
 KW biotin synthetase; biotin production; vitamin H; Biotin; ss.

OS Escherichia coli.

PN US5869719-A.

PD 09-FEB-1999.

PE 30-APR-1997: 97US-0846338.

PR 30-APR-1997: 97US-0846338.
 PR 08-MAR-1995: 95US-0401068.

* 125265 125364: gap of 100 bp
* 125365 141043: contig of 15679 bp in length
* 141044 141143: gap of 100 bp
* 141144 170896: contig of 29753 bp in length.

FEATURES

source

1. 170896

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="RP11-6P16"

/clone_lib="RP11 Human Male BAC"

1. 1744

misc_feature /note="assembly_fragment"

misc_feature 1845..2954

misc_feature /note="assembly_fragment"

misc_feature 3055..4839

misc_feature /note="assembly_fragment"

misc_feature 4940..7461

misc_feature /note="assembly_fragment"

misc_feature 7562..11912

misc_feature /note="assembly_fragment"

misc_feature 12013..15127

misc_feature /note="assembly_fragment"

misc_feature 15228..19176

misc_feature /note="assembly_fragment"

misc_feature 19277..24838

misc_feature /note="assembly_fragment"

misc_feature 24939..30632

misc_feature /note="assembly_fragment"

misc_feature 30733..37821

misc_feature /note="assembly_fragment"

misc_feature 37922..44638

misc_feature /note="assembly_fragment"

misc_feature 44739..49424

misc_feature /note="assembly_fragment"

misc_feature 49525..56517

misc_feature /note="assembly_fragment"

misc_feature 56618..65413

misc_feature /note="assembly_fragment"

misc_feature 65514..75659

misc_feature /note="assembly_fragment"

misc_feature 75760..86433

misc_feature /note="assembly_fragment"

misc_feature clone_end:SP6

misc_feature vector_side:left"

misc_feature 86534..98763

misc_feature /note="assembly_fragment"

misc_feature 98864..109919

misc_feature /note="assembly_fragment"

misc_feature 110020..125264

misc_feature /note="assembly_fragment"

misc_feature 125365..141043

misc_feature /note="assembly_fragment"

misc_feature 141144..170896

misc_feature /note="assembly_fragment"

BASE COUNT 41547 a 41714 c 41241 g 44377 t 2017 others

ORIGIN

Query Match

74.4%: Score 18.6; DB 2; Length 170896;

Best Local Similarity 84.0%; Pred. No. 4.2e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgaag 25

Db 95897 GGATCTGCTGTTCAATGAGCAGAG 95873

Search completed: December 26, 2001, 10:56:35
Job time: 1957 sec


```

/note="GRAIL"
/evidence=not_experimental
repeat_region
/rpt_family="LIPa3"
21815..21812
/note="AT_rich"
/rpt_family="Low_complexity"
21955..22142
/rpt_family="LIMD2"
complement(22044..22097)
/note="GRAIL"
/evidence=not_experimental
complement(22048..22097)
/note="MZF"
/evidence=not_experimental
22287..22436
/note="MZF"
/evidence=not_experimental
complement(22312..22414)
/rpt_family="FLAM_A"
22446..22486
/note="(CA)n"
/rpt_family="Simple_repeat"
complement(23834..23878)
/note="AT_rich"
/rpt_family="Low_complexity"
complement(23981..24687)
/rpt_family="LTR8"
25470..25541
/rpt_family="MER5B"
complement(25769..26077)
/rpt_family="AluSg"
complement(25979..26084)
/note="MZF"
/evidence=not_experimental
complement(26439..26499)
/note="(TA)n"
/rpt_family="Simple_repeat"
26526..26623
/note="MZF"
/evidence=not_experimental
complement(26982..27088)
/note="MZF"
/evidence=not_experimental
27646..27683
/note="MZF"
/evidence=not_experimental
complement(28316..28316)
/rpt_family="MLT1A2"
complement(28698..28721)
/note="AT_rich"
/rpt_family="Low_complexity"
complement(28989..29086)
/note="(TA)n"
/rpt_family="Simple_repeat"
complement(29452..29895)
/rpt_family="MER31B"
29515..29615
/note="GRAIL"
/evidence=not_experimental
29970..30011
/note="AT_rich"
/rpt_family="Low_complexity"
30034..30055
/note="AT_rich"
/rpt_family="Low_complexity"
30214..30544
/note="AT_rich"
/rpt_family="Low_complexity"
30223..30508
/rpt_family="AluSg"
30519..30539
/note="(TAA)n"

```

```

Query Match      74.4%  Score 18.6;  DB 9;  Length 151696;
Best Local Similarity 84.0%;  Pred. No. 41e+02;
Matches 21;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

OY      1  gacatgctgttgaagcagcagcag 25
          |||||
DB 125637  GCATCTGCTCTTTGGAAGCTGCAG 125613

RESULT  44
AL138781  LOCUS
          DEFINITION
          ACCESSION
          VERSION
          KEYWORDS
          SOURCE
          ORGANISM
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 154279)
          Direct Submission
          Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
          requests: clonerequests@sanger.ac.uk
          On Aug 9, 2001 this sequence version replaced gi:14530205.
          ----- Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          ----- Project Information
          Center project name: dj1079D1
          ----- Summary Statistics
          Assembly program: XGAP4, version 4.5
          Sequencing vector: M13; M77815; 14% of reads
          Sequencing vector: Plasmid; 108752; 85% of reads
          Chemistry: Dye-terminator Big Dye; 88% of reads
          Chemistry: Dye-primer-amersham; 11% of reads
          Consensus quality: 153247 bases at least Q30
          Consensus quality: 153539 bases at least Q20
          Insert size: 153679; sum-of-contrigs
          Quality coverage: 12.75x in Q20 bases; sum-of-contrigs
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contrigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contrigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          1      9651: contrig of 9651 bp in length
          *      9652 9751: gap of 100 bp
          *      9752 43191: contrig of 33440 bp in length
          *      43192 43291: gap of 100 bp
          *      43292 64482: contrig of 21191 bp in length
          *      64483 64582: gap of 100 bp
          *      64583 92209: contrig of 27627 bp in length
          *      92210 92309: gap of 100 bp
          *      92310 104336: contrig of 12027 bp in length
          *      104337 104436: gap of 100 bp
          *      104437 118874: contrig of 14438 bp in length

```


	/note="MZE"	/evidence=not_experimental
repeat_region	/complement(6630..6689)	/note="(TAA)n"
repeat_region	/rpt_family="Simple_repeat"	
repeat_region	6912..7214	/rpt_family="THEIC"
repeat_region	7241..8157	/rpt_family="THEI-INTERNAL"
exon	7605..7838	/note="GENSCAN"
exon	7822..7838	/note="XPOND"
repeat_region	/evidence=not_experimental	
repeat_region	8158..8538	/rpt_family="THEIC"
repeat_region	8703..8891	/rpt_family="LMC2"
repeat_region	8737..8891	/rpt_family="LMC2"
repeat_region	9211..9239	/note="AT_rich"
repeat_region	/rpt_family="Low_complexity	
repeat_region	10062..10156	/note="(CATAn)"
exon	/rpt_family="Simple_repeat"	
exon	10427..10531	/note="MZE"
repeat_region	/evidence=not_experimental	
repeat_region	10448..10943	/rpt_family="LTR2"
exon	11180..11252	/note="GRAIL"
exon	/evidence=not_experimental	
exon	11180..11252	/note="MZE"
exon	/evidence=not_experimental	
repeat_region	11180..11252	/note="GENSCAN"
repeat_region	12184..12573	/rpt_family="LTRA0a"
repeat_region	/complement(12677..12773)	
repeat_region	/rpt_family="L2a"	
repeat_region	12796..12875	/note="AT_rich"
exon	/rpt_family="Low_complexity"	
exon	/complement(13203..13221)	/note="XPOND"
repeat_region	/evidence=not_experimental	
exon	/complement(13313..13679)	
exon	/rpt_family="L2"	
exon	13549..13649	/note="GRAIL"
repeat_region	/evidence=not_experimental	
repeat_region	/complement(16076..16142)	
repeat_region	/rpt_family="MSA"	
repeat_region	/complement(16143..17744)	
repeat_region	/rpt_family="MST-INTERNAL"	
repeat_region	/complement(17745..18149)	
repeat_region	/rpt_family="MSA"	
exon	/complement(18237..18613)	
exon	/rpt_family="MSA"	
exon	/complement(20615..20739)	/note="GRAIL"
repeat_region	/evidence=not_experimental	
exon	20673..20725	/rpt_family="U2"
exon	/complement(20934..21068)	/note="MZE"
repeat_region	/evidence=not_experimental	
exon	20955..21248	/rpt_family="Aluub"
exon	21393..21437	

Center project name: 16593
Center clone name: 2017_B_4

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 737: contig of 737 bp in length
738 837: gap of 100 bp
838 1603: contig of 766 bp in length
1604 1703: gap of 100 bp
1704 2455: contig of 752 bp in length
2456 2555: gap of 100 bp
2556 3321: contig of 766 bp in length
3322 3421: gap of 100 bp
3422 4165: contig of 744 bp in length
4166 4265: gap of 100 bp
4266 5016: contig of 751 bp in length
5017 5116: gap of 100 bp
5117 5868: contig of 752 bp in length
5869 5968: gap of 100 bp
5969 6704: contig of 736 bp in length
6705 6804: gap of 100 bp
6805 7561: contig of 757 bp in length
7562 7661: gap of 100 bp
7662 8408: contig of 747 bp in length
8409 8508: gap of 100 bp
8509 9255: contig of 747 bp in length
9256 9355: gap of 100 bp
9356 10116: contig of 761 bp in length
10117 10216: gap of 100 bp
10217 10967: contig of 751 bp in length
10968 11067: gap of 100 bp
11068 11819: contig of 752 bp in length
11820 11919: gap of 100 bp
11920 12677: contig of 758 bp in length
12678 12777: gap of 100 bp
12778 13519: contig of 742 bp in length
13520 13619: gap of 100 bp
13620 14352: contig of 733 bp in length
14353 14452: gap of 100 bp
14453 15206: contig of 754 bp in length
15207 15306: gap of 100 bp
15307 16056: contig of 750 bp in length
16057 16156: gap of 100 bp
16157 16921: contig of 765 bp in length
16922 17021: gap of 100 bp
17022 17766: contig of 745 bp in length
17767 17866: gap of 100 bp
17867 18601: contig of 735 bp in length
18602 18701: gap of 100 bp
18702 19458: contig of 757 bp in length
19459 19558: gap of 100 bp
19559 20312: contig of 754 bp in length
20313 20412: gap of 100 bp
20413 21169: contig of 757 bp in length
21170 21269: gap of 100 bp
21270 22012: contig of 743 bp in length
22013 22112: gap of 100 bp
22113 22857: contig of 745 bp in length
22858 22957: gap of 100 bp
22959 23697: contig of 740 bp in length
23698 23797: gap of 100 bp
23798 24544: contig of 747 bp in length
24545 24644: gap of 100 bp
24645 25399: contig of 755 bp in length

25400 25499: gap of 100 bp
25500 26246: contig of 747 bp in length
26247 26346: gap of 100 bp
26347 27105: contig of 759 bp in length
27106 27205: gap of 100 bp
27206 27943: contig of 738 bp in length
27944 28043: gap of 100 bp
28044 28793: contig of 750 bp in length
28794 28893: gap of 100 bp
28894 29648: contig of 755 bp in length
29649 29748: gap of 100 bp
29749 30510: contig of 762 bp in length
30511 30610: gap of 100 bp
30611 31355: contig of 745 bp in length
31356 31455: gap of 100 bp
31456 32187: contig of 732 bp in length
32188 32287: gap of 100 bp
32288 33032: contig of 745 bp in length
33033 33132: gap of 100 bp
33133 33884: contig of 752 bp in length
33885 33984: gap of 100 bp
33985 34729: contig of 745 bp in length
34730 34829: gap of 100 bp
34830 35573: contig of 744 bp in length
35574 35673: gap of 100 bp
35674 36427: contig of 754 bp in length
36428 36527: gap of 100 bp
36528 37281: contig of 754 bp in length
37282 37381: gap of 100 bp
37382 38149: contig of 768 bp in length
38150 38249: gap of 100 bp
38250 39005: contig of 756 bp in length
39006 39105: gap of 100 bp
39106 39852: contig of 747 bp in length
39853 39952: gap of 100 bp
39953 40698: contig of 746 bp in length
40699 40798: gap of 100 bp
40799 41554: contig of 756 bp in length
41555 41654: gap of 100 bp
41655 42411: contig of 757 bp in length
42412 42511: gap of 100 bp
42512 43252: contig of 741 bp in length
43253 43352: gap of 100 bp
43353 44112: contig of 760 bp in length
44113 44212: gap of 100 bp
44213 44964: contig of 752 bp in length
44965 45064: gap of 100 bp
45065 45810: contig of 746 bp in length
45811 45910: gap of 100 bp
45911 46672: contig of 762 bp in length
46673 46772: gap of 100 bp
46773 47531: contig of 755 bp in length
47532 47631: gap of 100 bp
47632 48375: contig of 744 bp in length
48376 48475: gap of 100 bp
48476 49222: contig of 747 bp in length
49223 49322: gap of 100 bp
49323 50066: contig of 744 bp in length
50067 50166: gap of 100 bp
50167 50911: contig of 745 bp in length
50912 51011: gap of 100 bp
51012 51766: contig of 755 bp in length
51767 51866: gap of 100 bp
51867 52624: contig of 758 bp in length
52625 52724: gap of 100 bp
52725 53476: contig of 752 bp in length
53477 53576: gap of 100 bp
53577 54314: contig of 738 bp in length
54315 54414: gap of 100 bp
54415 55155: contig of 741 bp in length
55156 55255: gap of 100 bp
55256 56023: contig of 768 bp in length
56024 56123: gap of 100 bp

```

repeat_region      9118..9149
                    /note="8 copies 4 mer gaat 87% conserved"
misc_feature        11646..13304
                    /note="Cpg island"
                    /evidence=not_experimental
repeat_region      12015..12133
                    /note="7 copies 17 mer 65% conserved"
repeat_region      12408..12602
                    /note="13 copies 15 mer 59% conserved"
repeat_region      12430..12597
                    /note="56 copies 3 mer cgc 58% conserved"
repeat_region      12439..12614
                    /note="11 copies 16 mer 61% conserved"
repeat_region      12560..12659
                    /note="50 copies 2 mer cc 61% conserved"
repeat_region      12694..12779
                    /note="43 copies 2 mer gg 64% conserved"
repeat_region      13950..14036
                    /note="3 copies 29 mer 93% conserved"
repeat_region      14135..14458
                    /note="9 copies 34 mer 90% conserved"
misc_feature        complement(14531..14689)
                    /note="match: GSS: Em: AC074935"
repeat_region      14742..14917
                    /note="11 copies 16 mer 77% conserved"
repeat_region      15517..15804
                    /note="18 copies 16 mer 60% conserved"
repeat_region      15680..15730
                    /note="3 copies 17 mer 94% conserved"
misc_feature        complement(17738..18034)
                    /note="match: GSS: Em: B51354"
misc_feature        21357..22602
                    /note="Cpg island"
                    /evidence=not_experimental
repeat_region      21666..21857
                    /note="12 copies 16 mer 61% conserved"
repeat_region      21718..21853
                    /note="8 copies 17 mer 64% conserved"
repeat_region      21725..21894
                    /note="5 copies 34 mer 65% conserved"
misc_feature        join(21898..22001,22192..22592,23541..24088)
                    /gene="c380A1.2"
                    /note="match: ESTs: Em: AA045131 Em: AA213029 Em: AA310331
                    Em: A1723612 Em: A158791 Em: A1195565 Em: M91490 Em: A1674232
                    Em: A1015112 Em: A1015123 Em: AA335730 Em: AA643385 Em: C21054
                    Em: A1738487 Em: A1263152"
                    /evidence=not_experimental
mrna                join(21898..22001,22192..22592,23541..23600,23679..24088)
                    /gene="c380A1.2"
                    /note="match: ESTs: Em: AA158790"
                    /evidence=not_experimental
gene                /product="c380A1.2.1 (novel protein (isoform 1))"
                    /gene="c380A1.2"
                    /join(21898..22001,22192..22592,23541..23600,23679..23995)
                    /gene="c380A1.2"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="c380A1.2.1 (novel protein (isoform 1))"
                    /protein_id="CAB56188.1"
                    /db_xref="GI:5912460"
                    /translation="MGFPAALILCALCCGLAPAPAGYSEERCSSMRSGSLTQEPGSV
                    GQIALACAGNEMILYPACALRLITLGGPPRRAPGACIRPVPAPAGAOVAPRAGGA
                    LELLLEGGPCGRCVWGPERRRRLFLQATIRRHVSVAFFELREDCRPPLP
                    OAHGICVDGACRPSDAELLAACTSDPVIHGI HGVTHDVELQESVITIVAAVLRQ
                    TPLFOAGRSGDGLTSIRPLRCGVHPGPTFLFMGWSRFGSEARLGCAPRQERRRA
                    YEAARAHLPCCVALH"
                    join(21898..22001,22192..22592,23541..23647)
                    /gene="c380A1.2"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="c380A1.2.2 (novel protein (isoform 2))"

```

```

/protein_id="CAB56189.1"
/db_xref="GI:5912461"
/translation="MGFPAALILCALCCGLAPAPAGYSEERCSSMRSGSLTQEPGSV
GQIALACAGNEMILYPACALRLITLGGPPRRAPGACIRPVPAPAGAOVAPRAGGA
LELLLEGGPCGRCVWGPERRRRLFLQATIRRHVSVAFFELREDCRPPLP
OAHGICVDGACRPSDAELLAACTSDPVIHGI HGVTHDVELQESVITIVAAVLRQ
TPLFOAGRSGDGLTSIRPLRCGVHPGPTFLFMGWSRFGSEARLGCAPRQERRRA
YEAARAHLPCCVALH"
complement(23770..24065)
/note="match: STS: Em: R68197"
24088
/gene="c380A1.2"

BASE COUNT      4325 a      8228 c      7960 g      4529 t
ORIGIN
polyA-site
misc_feature
polyA-site
Query Match      74.4% Score 18.6; DB 9; Length 25042;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1      ggaatcgtgttgaagcagcagcag 25
Db      23196      GGAATCTGAGTCTCAAGACGACGAG 23172

RESULT 41
AC068957/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 70326)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 21, clone CTD-2017B4
Unpublished
2 (bases 1 to 70326)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bede,F.,
Boguslavsky,I., Bouckgatter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,D.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McKean,P., McGurk,A., McKernan,K., McPheters,R.,
Mceldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirelli,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

```

promoter          EVAHEDKGAIGADGTNIEFAKELARHFIDIDLT"
                  9705..9710
                  /note="promoter 2"
mat_peptide       9734..10159
                  /gene="RNA polymerase subunit operon"
                  /product="ribosomal protein Hns12"
                  9734..10162
                  /gene="RNA polymerase subunit operon"
                  9734..10162
                  /gene="RNA polymerase subunit operon"
                  /codon_start=1
                  /transl_table=11
                  /product="ribosomal protein Hns12"
                  /protein_id="CAA0429.1"
                  /db_xref="GI:43545"
                  /db_xref="SWISS-PROT:P15756"
                  /translation="MTNGKVAARKLKDRORMSDEYARRERGLCKKSDPLEGAPQ
                  GRTVLEKVGCEAKQPSNLRKRVOLINKKQVAFCEPDGAISFIDEHDEVTIAG
                  IGGAAGKRAKGLDSCVNRKVEKVCNLSLIEVRGNAKPVK"
                  9734..>10798
                  /gene="RNA polymerase subunit operon"
                  10166..10798
                  /gene="RNA polymerase subunit operon"
                  10166..10798
                  /gene="RNA polymerase subunit operon"
                  /codon_start=1
                  /transl_table=11
                  /product="ribosomal protein Hns7"
                  /protein_id="CAA0430.1"
                  /db_xref="GI:43546"
                  /db_xref="SWISS-PROT:P15763"
                  /translation="MSDEQPADEDETEAAESEDQEVANAKLFEKRWDAELIYEDP
                  STRRYLVATPVVAHMGHRAOKRKSEISIVERLANRLMKGNACKKQQAALTKVDA
                  FDIYHEKTDENPIQVIVSAVENAAPRETVRLKYGISVPQAVDTAPQRRVDQALKFL
                  ADAHSAISFKTPIDDAEALANQLAGADYVWQYIAIGOKKEKRRVAAAR"
                  10166..10795
                  /gene="RNA polymerase subunit operon"
                  /product="ribosomal protein Hns7"
BASE COUNT       2218 a 3453 c 3887 g 1650 t
ORIGIN
Query Match      74.4%: Score 18.6; DB 1; Length 11208;
Best Local Similarity 84.0%: Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 gqatctgctgttgaagcgacag 25
Db 6721 GGATCTGCTCTTGAGCGCGCATCAG 6697
RESULT 40
LOCUS            HS380A1 25042 bp DNA PRI 19-APR-2001
DEFINITION       Human DNA sequence from clone LA16-380A1 on chromosome 16. Contains
                  two novel genes, ESTs, an STS, GSSs and five putative CpG islands,
                  complete sequence.
ACCESSION        297653
VERSION          297653.14 GI:13929477
KEYWORDS         HTG; CpG island; human.
SOURCE           Homo sapiens
ORGANISM         Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE        1 (bases 1 to 25042)
AUTHORS         Wallis, J.
JOURNAL          Direct Submission
                  Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
                  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
                  requests: clonerequest@sanger.ac.uk
                  On May 2, 2001 this sequence version replaced gi:13277290.
                  During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
COMMENT

```

FEATURES

source

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/WormPep LA16-380A1 is part of a clone contig from the tip of the short arm of chromosome 16 spanning 2Mb of p13.3 (Higgs D.R., Flint J., Daniels R., MRC Molecular Haematology Unit, Institute of Molecular Medicine, Oxford (unpublished)), and is from the Los Alamos, flow sorted human chromosome 16 libraries constructed by Norman Doggett (unpublished). VECTOR: scos-1 IMPORTRAM: This sequence is not the entire insert of clone LA16-380A1. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone LA16-444C9 is at 24943 in this sequence. The true right end of clone LA16-313D11 is at 100 in this sequence.

```

misc_feature      1..25042
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="16"
                  /clone="LA16-380A1"
                  /clone_1kb="LA16"
                  533..2551
                  /note="CpG island"
                  /evidence=not_experimental
                  1892..1999
                  /note="54 copies 2 mer cc 63% conserved"
                  1913..2012
                  /note="20 copies 5 mer ccccg 78% conserved"
                  1918..2022
                  /note="7 copies 15 mer 76% conserved"
                  2044..2433
                  /note="5 copies 78 mer 67% conserved"
                  complement(join(2188..2531,3381..3680))
                  /gene="c380A1.1"
                  complement(join(<2188..2531,3381..>3680))
                  /gene="c380A1.1"
                  /note="this gene fragment and fragment c380A1.1a are part
                  of one gene; GENSCAN suggests the second exon of this
                  fragment and the first of fragment a are the ends of one
                  large exon
                  match: ESTs: Em:H49535 Em:H49776"
                  /codon_start=1
                  /evidence=not_experimental
                  /product="c380A1.1 (novel protein)"
                  /protein_id="CAB56187.1"
                  /db_xref="GI:5912459"
                  /translation="CEKCVLAQVCKARRRVLQPKFVAGLTPLVIAKELYNVLPGEK
                  EFVNIQGFARKEFGFCIVGSDIDCEFDNALSKKGVKAMSLKSTTTDGLFVMA
                  LEQMGVVRLELSGNDFTAGLWSSLSARTISLSVDCINVDADAIAAISLPPLPA
                  ELISQAHVVDFTALAYFTARGHSTHRLRLSCWEITNHGVVNVVHSLPLRFL"
                  2832..2871
                  /note="10 copies 4 mer ttta 87% conserved"
                  3482..4025
                  /note="CpG island"
                  /evidence=not_experimental
                  5552..5587
                  /note="18 copies 2 mer aa 83% conserved"
                  7007..8059
                  /note="CpG island"
                  /evidence=not_experimental

```

AUTHORS Jeffers, H., Gropp, F., Lottspeich, F., Zillig, W. and Garrett, R.A.
 TITLE Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles Halobacterium halobium and Halococcus morrhuae
 JOURNAL J. Mol. Biol. 206 (1), 1-17 (1989)
 MEDLINE 89199633
 REFERENCE 2 (bases 1 to 11208)
 AUTHORS Jeffers, H.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-1990) Jeffers H., University of Aarhus, Institute of Medical Biochemistry, Universitetsparken Bygn. 170, DK-8000, Aarhus C, Denmark
 COMMENT See also <X57145>
 FEATURES Data kindly reviewed (23-JAN-1991) by Jeffers H.
 source location/Qualifiers
 1. 11208
 /organism="Halobacterium salinarum"
 /strain="R1"
 /db_xref="taxon:2242"
 /clone_lib="Lambda EMBL3"
 /note="promoter 1"
 promoter /note="promoter 1"
 1380. .9570
 /gene="RNA polymerase subunit operon"
 join(1380. .9570, 9734. .10798)
 /gene="RNA polymerase subunit operon"
 1423. 1650
 /gene="ORF 75"
 1423. 1650
 /gene="ORF 75"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA40423.1"
 /db_xref="GI:43539"
 /db_xref="SWISS-PROT:P15740"
 /translation="MVDVQHELVPEHTVLEDETLGVIAVDYIDRTLPKIKYKQPA
 LPDNEIGDVEIVRDSRTDEAVYRLVIE"
 1423. 1647
 /gene="ORF 75"
 1651. 3222
 /gene="RNA polymerase subunit operon"
 1651. 3222
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit B''"
 /protein_id="CAA40424.1"
 /db_xref="GI:43540"
 /db_xref="SWISS-PROT:P15352"
 /translation="MLQEDKRELSESYFSKRLAEHFRSYNAFLHGQMDVYTER
 IETDIDGKDEDEPVWVELGDAVPTPRVREADGSEELIYPOEARLNTITYSAVPEM
 STYRGGEERERLYDTTETKVGKMPIMVGSCKNTSGSDEELIEIGEPAPGPGFI
 INSERVLMTSEDLAKILAEYDSKYGGEIPVSDPEITVAFLENLBEAVQOEALIE
 VSPFVSQISPTVILVRAKLSDEDEIVHRSDEPEITVAFLENLBEAVQOEALIE
 DLGQVAVSGQKNYQKLRANYIDRYLPLHLEHVEEETRIKNAVYLCRAEACFE
 IALGREADDKKHAYANKRLKVSGLDKMLDLFTALNKLAIDVAYYOLERAMRKRELTN
 TVYRSDVILTERLEHPIATGNWVGSRGVSQIOLVDTDFMGVLSHLRLRSLRSQPHF
 EARDLHATQWRICPSETPEPGKNGIYKVFQAQAMLSQDVDERDLKQLSMGVEGI
 GISMETSTSTASDD"
 1651. 3219
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit B''"
 3224. 5047
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit B''"
 3224. 9151
 /gene="RNA polymerase subunit operon"
 3224. 5050
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit B''"
 /protein_id="CAA40425.1"

mat_peptide /db_xref="GI:43541"
 /translation="MSTEEAAVYNGSLVGNHNEPEPLADQIREARRRGVSEWVY
 SVADRGEIVVADAGRARPLLVENEGPVYTOGEVAVAKNGIDFEDDVAEAGVER
 IDAEEDDILVNADEEELTDTHTLEIDQLLGIGAAMIPPEHNASPRITMGAGM
 KOSLGLPANNYLRIPDTRQHLHYPOKAMVNTQTEIGYDPRPAGQNFVAVMSGEY
 FNIEDALVNKSGSERVALSRSHFEPTEGEERARYGCOEDREIIPGDVFAVAYDAY
 THLDDGLVNPETKVDSSVLGKTSPPFLPEEDMGSLSPKRETSVYMSGEED
 VVDITVLMGEDGSKIAKSVYDERIPELGDKFAPARHOKGVYGLAPOEMPFQEG
 VEDVIALNHALPSRMTVGHVLEMGKAGSDGSDGTPPTGSEDEIRTELEDRG
 FKSGKEVNYSGVSGEIEAEIEFVGTIFPKLHYMSNKLARSGVQVQLTRQTEG
 RAREGGLRVGEMRDVIGHGAAMVLEKRLDSDREELHVGNGMGMAVENYEQRY
 YCNCEETDVHSTIEMSAFKLLIDEMKALGIAPRLEAAV"
 5050. 7965
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit A"
 /protein_id="CAA40426.1"
 /db_xref="GI:43542"
 /db_xref="SWISS-PROT:P15350"
 /translation="MSAGAPKEIGELISGLMDPEYRDMSATKYITADTYDDGEPRI
 DMGIMDPRLVGIDPLECKTGQSGSGCGHFEHTELAPVIAHGVSKLIRLKETC
 RECASLILTEEEKDEYRENDLRSLRQDVDMVAIIEARKKDCPHCGVQYDVK
 HEKPTYYEVOQVLAISDYSERILASMQPDEDDAGVSQELAEQTDISRINETLS
 GERPREDEBALETAIGADLTPEDMNKIMPSSIIMDFDITPEDEALGAVNSDRP
 EMMLTVLPVPVPTARPSITLDGQSEDDLTHTKLVDTIRINQRFMEKEACAPOLII
 EDLIMELQYVHTTFMNEISGTPPARHSRQPLKTSQRLKREKFRSGISCKRRNF
 SARVILSDPTLLNEGYDPRVATEMTQMVVNEONLRAKRYVYNGEGHGANTV
 TRDGRVRYTEKVECELAERVPGWEVQHHLDGDIIFENRQPSLHRSIMAHVVV
 MPKPTPLNTVCPVYVADFDGDEMNNHLEAENEAERVLARVOEQLISPFENGI
 ICAIDHISTYLLITNDNPRENTOASDILRQTRIDELPAAGTDEGQYVHCNTPAR
 SELLPDLSLEFTGTGDIYIDGQLLBSTIADDEYGEYSITVPTITTKVHCNTPAR
 IFINEVASLMSRISMEHGFSGIDIDEFTSEAREIDEAIOAIDVQELIETENGDI
 LESLPGRVDETLEKIMQVTLGKARSDAGVAENEDENPVAVMANSARSGSMLNT
 OMACQVCOAVRGELIRNGYEDRTLSHFAPNDLSSBAHGFEVNSYSGTLKPEFFHA
 MCGREGIVDTAVRTSRGLQRLINLSELLEYOYGVYDITDITVOEFGFGDQSP
 VOYSSNEVDIVDEHVAADRLINSEFSDTKAFLIEVEPPTMLSEGAAMEVESD"
 5050. 7962
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit A"
 7958. 9148
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit C"
 7958. 9151
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit C"
 /protein_id="CAA40427.1"
 /db_xref="GI:43543"
 /db_xref="SWISS-PROT:P15354"
 /translation="MTEFDYDVGHVYIDETELPRRLKDEVLATAEERGVTKQANET
 ATAVEAOYLDRPDLDPVGVSAOSIGCEPTQMTNTPFHAGVAENDVQOGLPRLE
 LVDAKRPDPPRVMPVYLEDYAEERBAHEVVMTEATKILAGDITSTNADAVYRDI
 LNEEDLQRMVRVNSTQJAKVAVETIEGNLGVYVPODGTITTEGSPSPYRELQVY
 EQLDILYFKGILEEYSRAVITRETEERDEBEVLYTIEGSAFKKALKIEVDATRTSCNNT
 HEVHTLGTIEAAREAIINETMTTLEEDGLDVNIRHMLVAIDLINTNGTIDESIGRHI
 SGNKTSYLARAFAEYTVNHLDAAIHESDDLDGVIENTVYGVKRVRLGTGVDYLRMGA
 TQTSND"
 9151. 9567
 /gene="ORF 139"
 9151. 9570
 /gene="ORF 139"
 9151. 9570
 /gene="ORF 139"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA40428.1"
 /db_xref="GI:43544"
 /translation="MTITLSDBARQLIAVFTETIDATAVDCLPDDDRVYVYIAGENG
 AATIDGGSRVDALEATITLGRSVYLVEDATFAGCVFANMLSPRAVYNNVYSENDITVAVA

```

/product="30S ribosomal protein S12P"
/protein_id="AAG20689.1"
/db_xref="GI:10582040"
/translation="MTNGKVAARKLKRDQORMSDSFYARREGGLCKSPLEGCARQ
GRCIYLKVEGTAKOPNATIRKCYRVOILKNGKVTATCPEDGATISFLIDEHVTIAG
IGGAGRAMGDLGCVNTKVERKNGSVSLLEIRGNKAFVR"
complement(2263..2682)
/gene="nusa"
/note="VNC2661G"
complement(2263..2682)
/gene="nusa"
/note="nusa"
/codon_start=1
/transl_table=11
/product="transcription termination antitermination
factor"
/protein_id="AAG20690.1"
/db_xref="GI:10582041"
/translation="MTIRLSDEARQLIAVETETDATTAVDCLPDDRVVVVAVGEMG
AAGDGSRDVALEATLGRSVLVEDAPTAEGFVANAISPAAVYNTVSENDPTVAVG
EVAHEDKGVAIGADGTNIETAKELAAHFIDIDILQI"
complement(2682..3875)
/gene="rpoC"
/note="VNC2662G"
complement(2682..3875)
/gene="rpoC"
/note="rpoC"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase subunit C"
/protein_id="AAG20691.1"
/db_xref="GI:10582042"
/translation="MTFEDVTDVGHVYIETDELPRRLKDEVLAIEERGVTKSQANET
ATAEAQYLDTRVDLPDVGTVSAOSIGEPGOMTNTFHYAGVAEMDVYOGCPRIIE
LVDAKTPDTPVMEVLEDEAEEREAHVVKITATKILALGISTNVADWAVRDL
LNEDELOERMPRVSTQIAGEVAETTEGMLGVYVODGRTILEGSPSEPSRKLLOLV
EOLBDIVFKGIEEESRVIRIREEEREEVLTTEGSAFKALKIEGVDFRTSCNLT
EYHKTIGTIAAREALINETNTTILEGGLDVNIRHLMVADIMTNDGTIESIGRGI
SGKNSYLAAREAEVYVNLHLLDAIHGSDLDLVGLENVTGKPVRLGTGVDLRMGA
TQTSID"
complement(3868..6648)
/gene="rpoA"
/note="VNC2664G"
complement(3868..6648)
/gene="rpoA"
/note="rpoA"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase subunit A"
/protein_id="AAG20692.1"
/db_xref="GI:10582043"
/translation="MGLMDPRLDVTDPLGECTKCGQRSGGCGNGFHGHIETAPVIVHC
FSKILRLRGTCRECAISLLIEEKDEYRENDTRSLKODSDVMTAIRARKD
HCHPGCEVOYDAKHEKPTTYEVOQVLAISYSERIAASHQPDDEDDAGVSPLELQ
TJDDISRIINELSGEFPREDREALETAGADLTLENNKMLPSDIRMPEDIPED
LEALGVNDSRDEMMLITVLPVPTVATRASTLDPGOSSEDDIPIKIVDIIRINRF
MENREAGAPOLITIEDIMELLOYHTVPMNEISGTPARRRSRLPRTLSQKAGEG
RPFSGISGRKVNYSATVISPDTLSINEGVDPDRAVTEHTQTMVNEONLEAKRGV
RNPBEGHGANVTRPDGRVRVTEKVCPEELARVAPLHLEIDGDIIFNROPS
LHRMSIMAEVNVVMPYKTERLNTVCPVNAFDGDEMMHALLNEARAEARVLRV
QEOILSPRESENTIGAIOHISGTYLTNDNPFNETQASDLRQTRIDELPAACTO
EDGDQVYGHQIFSELDPDLSEFTGTGDTVLEIDGOLLEGTIADDEGVGSGRTV
DITIKVHGNTRARIFINEVASIAMSIMHGFSGITGIDDEFVSEAREPIDEAIOQSYD
ANGCAGSMILNQMGACVGOCAVGERINRGEDTLSHPAVNDLSSAHGFENSVY
TSGLTIRKEFFTHMAGREGVLVDIATKTSGLYLRULIALSELETQYGTAVQTSDT
TVQFERGEDGTSFVQVSSNEEVDIDVEHADRILNFEPSDIOKAEFLVEEPPITLS
EHGAAMEVESDD"
complement(6783..8609)
/gene="rpoB"
/note="VNC2665G"
complement(6783..8609)
/gene="rpoB"

```

```

/note="RpoB"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase subunit B"
/protein_id="AAG20693.1"
/db_xref="GI:10582044"
/translation="MSREERAKVYVNGSLVGTNENPEELAEQIREARRRGVSEWNV
SVDRPGEVIVNADACRARRPLVYNGEPVYTOEEVAVKNGDDIDEDLVAGKVEF
IDAEEDDILVGVEEELTTDTHLEIDQLIFGIGAGMI PYPENASPRITMGAMM
KOSLGLPAAVYRI RPDTRQHLHYPOKAMVNTQTEQOIGDPRPAGQNVVAVMSYEG
FNTEDALVNMKGVSERALSRSHEFTYEGERERYPGQDRPEITPGDVRGARBDAV
THLDDGLVMEETKVDSSVLGKTSPPRELEPEDMGGLSPKRRRESYMRSSEDC
VVDVTVMGSEDSKILAKVSVDERIPELDGKPAASHCKGKVYGHIALAQEDMPFQEG
VPPDLVNHALPSSRTVGHVLEMLGKAGSIDGNSVDGTPPTGEGEDIRKTELDRG
FKSSGKLVMSGVSGEKIEAEIPEVGTIFYKLIYHWSNKLARSKGPVOLLTRPTEG
KAREGGLRVGEMERDVTIGHGAMVLEKRLDSSPDEELHVCNGCMATVENYEQRRV
YCPNCEETDVHSIEMSYAFKLLDDEMKALGIAPRIIEAV"
complement(8611..10182)
/gene="rpoB"
/note="VNC2666G"
complement(8611..10182)
/gene="rpoB"
/note="rpoB"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase subunit B"
/protein_id="AAG20694.1"
/db_xref="GI:10582045"
/translation="MLQEDKRELSESYSEKRLAEHHRFSYNAFLHGMQDVTEKER
IETIDGDKDEEPVWVLEGDVRAVTPRVAEAGSEELLYPOEARLNTTYSAPVEMM
SIVRGEEEREEVLTTEETKVGKMPIMVGSQDKNTSGFSEDELTIGEDPADGGYFI
INGSERVLMSTEDLAPNKLIAFYDSKYGDIOVATKPSORRCRYALVYENRBILLE
VSPRYSISISVTVLRALGIDSEDEIYHRSDEPIYKFMLENEADVOTQEBATE
DLCORASQCKNTOLKRAVYIDRYLPLHLEDVEEETRINAAVLYCRMAACEF
LALGREADQKHVYANKRLKVSQDKMDLFYALKLARVYKQYOLEARNMNRRLTVN
TVVRSDVLTRELEHP IATGNMVGCGSGLVQLVDRDFMGVLSHLRLSPLSRQPHF
EARDLHATQWLRICPSETPEGPNCGLVKNQAAMELSQVDVDERDLKQELSMGVEGI
PGISMETTSTTSADD"
complement(10183..10410)
/gene="rpoH"
/note="VNC2668G"
complement(10183..10410)
/gene="rpoH"
/note="rpoH"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase subunit H"
/protein_id="AAG20695.1"
/db_xref="GI:10582046"

```

Query Match 74.4% Score 18.6; DB 1; Length 10529;
 Best Local Similarity 84.0%; Pred. No. 3.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtcgtttgaagcgaacg 25
 ||||| ||| || ||| |||
 Db 5112 GGATYCGCTCTTGACGACGCATCAG 5136

RESULT 39
 HHRNAPOP/C
 LOCUS HHRNAPOP 11208 bp DNA BCT 06-JUN-1991
 DEFINITION H. halobium RNA polymerase subunit operon.
 ACCSSION X57144.X14999
 VERSION X57144.1 GI:43538
 KEYWORDS ribosomal protein; ribosomal protein S12; ribosomal protein S7; RNA
 polymerase; RNA polymerase subunit A; RNA polymerase subunit B;;
 RNA polymerase subunit C.
 SOURCE Halobacterium salinarum.
 ORGANISM Halobacterium salinarum
 Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 Halobacterium.
 REFERENCE 1 (bases 1 to 11208)

BASE COUNT	555	a	237	c	418	g	400	t
ORIGIN	<p> RPNFLINDAOSD, SAADIDYIIOGARSALTEGEMINGESI KCAVILLINWGAANAPDPISPGEHAIYINRLAEI EKFEODBRFPQONKEK EIDEYGEELRS IROGKCM I SLEESGEMEMETIKSI KGMEME IELKHTNDLOTLRGLTKEASMRNORBEELI SYTNMNRVYLODA I SVNOELHEEAVG ELDMGAEVLETTAEVYPVLCAGAA NLA SVRAEVAKIKLKEVI OKLSGVDLSMTYK SLIOGRKMSILLRRNADGDVIOEKDLELVONDI LALVREIGPERRHLEENI VPKI LEXKM DKHOKPKYVAHAKM INKFEFHKR IITXYAPMDSYVEMFKCAVPHHKEGFGTDLPI DEVFEEDLKVEAHOLIEGATFVGRSFRQI RDPFFYANMNSGASEIHKRKLQSSSA HPTLYGSDYDIOISTDLYIAHANOLVTNEDIOHLVRCFLHFORPTIMALLHGEVIMT RPNFLINDAOSD, </p>							
Query Match	74.4%: Score 18.6; DB 14; Length 1610;							
Best Local Similarity	84.0%: Pred. No. 2.6e+02;							
Matches	21;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps
Oy	<p> 1 ggaatctgctgtttaaagcgcagcag 25 </p>							
Db	<p> 1214 GGAATCTGAAGGTTGACGCGCATCCAG 1238 </p>							

RESULT	35				
LOCUS	AF188738S2/c	AF188738S2	2191 bp	DNA	PRI
DEFINITION		Homo sapiens LHX3 gene, Intron 1b.			16-Apr-2000
ACCESSION	AF188739				
VERSION	AF188739.1	GI:7576718			
KEYWORDS	.				
SOURCE	2 of 6				
ORGANISM	human.				
REFERENCE		Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 2191)				
JOURNAL	Sloop,K.W., Showalter,A.D., Von Kap-Herr,C., Peltanai,M.J. and				
MEDLINE	Analysis of the human LHX3 neuroendocrine transcription factor gene				
PUBMED	Gene 245 (2), 237-243 (2000)				
REFERENCE	20183685				
AUTHORS	10717474				
TITLE	2 (bases 1 to 2191)				
JOURNAL	Sloop,K.W., Showalter,A.D. and Rhodes,S.J.				
MEDLINE	Direct Submission				
PUBMED	Submitted (22-SEP-1999) Biology, IUPUI, 723 West Michigan Street,				
REFERENCE	Indianapolis, IN 46202-5132, USA				
AUTHORS	Location/Qualifiers				
TITLE	1..2191				
JOURNAL	/organism="Homo sapiens"				
MEDLINE	/db_xref="taxon:9606"				
PUBMED	/chromosome="9"				
REFERENCE	/map="9q34.3"				
AUTHORS	1..2191				
TITLE	/gene="LHX3"				
JOURNAL	/number=1b				
MEDLINE	354 a	647 c	783 g	406 t	1 others
PUBMED	BASE COUNT				
REFERENCE	ORIGIN				
AUTHORS					
TITLE	Query Match	74.48:	Score 18.6:	DB 9:	Length 2191:
JOURNAL	Best Local Similarity	84.0%:	Pred. NO. 2.7e+02:		
MEDLINE	Matches 21: Conservative	0:	Mismatches 4:	Indels 0:	Gaps 0:
PUBMED	1 ggaatgctgttttgaagcgagcagc 25				
REFERENCE					
AUTHORS	Db 1836 GGATCTGCTCTCTGTGAGCCCGCAG 1812				
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					

ACCESSION	complete cds.
VERSION	M69036
KEYWORDS	pbph gene: GI:141961
SOURCE	Alcaligenes eutrophus: protein H: protein I.
ORGANISM	Alcaligenes eutrophus (strain H16) DNA.
REFERENCE	Ralstonia eutropha
AUTHORS	Bacteria: Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
TITLE	1 (bases 1 to 4501)
JOURNAL	Pris,A., Priefert,H., Kruger,N. and Steinbuechel,A.
MEDLINE	Identification and characterization of two Alcaligenes eutrophus
FEATURES	gene loci relevant to the poly(-hydroxybutyric acid)-leaky
source	phenotype which exhibit homology to psh and pshI of Escherichia coli
	J. Bacteriol. 173, 5843-5853 (1991)
	91358378
	Location/Qualifiers
	1..4501

misc_feature	750	/gene="phb" /note="transcription start site" /evidence=experimental 1052..1321 /gene="phbi" 1052..1321 /gene="phbh" /codon_start=1 /transl_table=11 /evidence=experimental /product="protein H" /protein_id="AAI21977.1" /db_xref="GI:141962" /translation="MGDRPTIIINKIGLHARASAKLTQAGNFVSQYKMSRNGQVDA KSINGVMVLAMGICSTVYLETDGPDEQEMDLALLIANREGE" 1371..3149 /gene="phbi" 1371..3149 /gene="phbi" /codon_start=1 /transl_table=11 /evidence=experimental /product="protein I" /protein_id="AAI21978.1" /db_xref="GI:141963" /translation="MPFALHGIPVSRGAIVGRAHLILAPALDHSVHYLVDEDOILAAYEVE RLRAARAVARAEIALAKROLPRDPAPEELGAFIDYUMLITLDALAREPGLIRGRNN AEMALTTRLEIMHQDEITEDYLREKRDIQROYERTIKALGAPVLVPAPVALA DGEATGVIVAHDIAPADMLORHTVFHFEGVTDMGRTSHAIIVASLDIIPAVQV SASELIRODWIIIDDGAGLVIVDPPIIILEEVRIHQSRRALEKKRLORIHPAVVL DGLEIDLILANITEMADGALAAGAAGVGILFSEFLNMRDELCEDEQOACRGAV DAMEGLPVTTITIDIGADKPLDAGEFEETALNPALGLAIRWSLSEFGMTQLRAL LRASFQGVRLVLMPLAHASEIIDOTLAILAKKROLDERGEAYDPGMVGMTIEPPA VLLIFLPFRKMDFLSICTNDLTOTLALDRADNAVAHI.FDPLHPAVLDVARTIREAN RAGPVAVCGEMAGDPSMTRLGMCGLREFSMHPOLRVKOEILLHACELEPLVDO VLQADPEQARPLRWHPDEALIVE" 3156..3209 /gene="phb" 3156..3209 /gene="phb" /note="potential terminator"
BASE COUNT	760 a 1529 c 1483 g 729 t	
ORIGIN		
Query Match	74.4%:	Score 18.6; DB 1: Length 4501;
Best local Similarity	84.0%:	Pred. No. 2 ge+02;
Matches	21: Conservative	0: Mismatches 4: Indels 0: Gaps 0:
xy	1 gaatctgctgtttgaagcgacgcagc	25

[illegible]

```

/codon_start=1
/db_xref="FLYBASE:Frano0003985"
/db_xref="FlyBase:FBgn0004575"
/feature_id="AAFE54506.1"
/db_xref="GI:7299312"
/translation="MPPPPADGAGAAPELISFPGAKTPATAAPAPPGVSAPTSE
AKRESGLIAROVSLIGAAPDQAGSLICAAVQSMTOGAPEFSKDKFTLLVLIDONTW
SKYFERGRHLHGFDRIEVAQAEFRDIIVASASADIPGVYMAARSGTRARSRPDEYV
IROPDGGSDSVRFTLLICAGVPSINSTLHLYXODCPWFMSHLLDLOLRSLDGE
PLIEQFPFNPSHDIISRGMMRWLHVLLHSCFSLSMILDHOVAITICPVPYPPH
RSTRCTPEPVLHIQFOPTKEPSPVHLKAGHGVATRLNENSLAAGVIVAGNAGN
SVCYCTTEPYIDAKTSVHIQKTNKKYKAPRKSTITGMKNTNGOSAMLEQITLTKYKS
WVCEISELGEAGVEGVCISVVAKDGREYIISACDSTFALIGTQEDRQDLIDLVSQ
MOWCEPGRSNAQOEGPKILPSRSVSRAESEPILDGVAFTPLIAGBPAPAGMGPPI
RTSPAGVSGIQLRSLSKSIISEVEPESSGSGYVGGELJANKRKRAEFSGSGSGY
PSAGSGSGESPLKQOFSEVAPGEGVITGTOPLORSEEPALPTTASARAPSSY
SVSDSNTNQTILIRAGAGVQVYTNFQOEYVFNDEPKSGSAAHSIHSSSSITSSS
SISSTRINRGNALIOSPPAGGPPPPPPNTAVGNASAGGCTPMSFSSLSKQTSY
GNGSTSTVEITITRDNTNTNIGATATEGAGCVATIRINISNGIYAPTTGTTTYS
VTTINDMRSLAIGRMSASVYSAAPAATVLPDGTGSOYNSIASJGEBGLNMPDPSYR
PYSYRESNSAKSHDLDAVFDGRTTPASVGNKRYTRAGSISIDAMITGPPSNYKTY
DRGAKRSKMSMTSGSGVSGNGSGSLGQYIYPSIONAASFDPSSGMSSTIGSHTKK
WASACEDELDLK"
complement(join(43382..42513,42653..43196,43255..43473,
43581..43666,43738..43870,51247)..51476))
/gene="Utmp"

```

Query Match	75.2%	Score 18.8;	DB 3;	Length 221888;
Best Local Similarity	90.9%	Pred. No. 3.4e+02;		
Matches	20;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0.
OY	2	gattctgctgttgaagcgacg	23	
db	178558	GATCTGCTCTTGAAGCGCAGC	178537	

RESULT	34
LOCUS	AB014726
DEFINITION	AB014726 1610 bp RNA
ACCESSION	AB014726
VERSION	AB014726.1 GI:3986142
KEYWORDS	outer capsid protein VP5.
SOURCE	Chuzan virus CDNA to genomic RNA.
ORGANISM	Chuzan virus
REFERENCE	Viruses; dsRNA viruses; Reoviridae; Orbivirus; Palym virus group.
AUTHORS	1 (sites)
TITLE	Yamakawa,M., Furuchih,S. and Minobe,Y. Molecular characterization of double-stranded RNA segments encoding the major capsid proteins of a Palym serogroup orbivirus that caused an epizootic of congenital abnormalities in cattle
JOURNAL	J. Gen. Virol. 80 (Pt 1), 205-208 (1999)
MEDLINE	99131402
REFERENCE	2 (bases 1 to 1610)
AUTHORS	Yamakawa,M.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAR-1998) to the DDBJ/EMBL/Genbank databases. Makoto Yamakawa, National Institute of Animal Health, Department of Exotic Diseases, Joshiincho 6-20-1, Kodaira, Tokyo 187-0022, Japan (E-mail:yamakawed@affrc.go.jp, Tel:+81-42-321-1441, Fax:+81-42-325-5122)
FEATURES	Location/Qualifiers
SOURCE	1..1610

```

CDS      21..1586
         /codon_start=-1
         /product="outer capsid protein VP5"
         /protein_id="BAA34934.1"
         /db_xref="GI:3986143"
         /translation="MGRFVKFLKRAKGTITMALISNFAKSIITYGVKTIQKVAESEL"

```

```

misc_feature /clone="Rp11-306E15"
              /clone.lib="RP11-11.2"
              1..7271
misc_feature /note="assembly_fragment:00252"
              /372..26649
misc_feature /note="assembly_fragment:00307"
              /26750..30691
misc_feature /note="assembly_fragment:00433"
              /30792..59813
misc_feature /note="assembly_fragment:00758"
              /59914..62315
misc_feature /note="assembly_fragment:00792"
              /62416..78769
misc_feature /note="assembly_fragment:01028"
              /78870..90625
misc_feature /note="assembly_fragment:01048"
              /90726..96694
misc_feature /note="assembly_fragment:01089"
              /96795..105609
misc_feature /note="assembly_fragment:01119"
              /105710..115711
misc_feature /note="assembly_fragment:01128"
              /115812..163536
misc_feature /note="assembly_fragment:01253"
              /163637..171008
misc_feature /note="assembly_fragment:01502"
              /171109..177629
misc_feature /note="assembly_fragment:01571"
              /177730..193594
misc_feature /note="assembly_fragment:01659"
              /193595..36754
BASE COUNT 60811 a 36754 c 35693 g 59030 t 1306 others
ORIGIN

```

```

Query Match 75.2% Score 18.8 DB 2: Length 193594:
Best Local Similarity 90.9% Pred. No. 3.4e+02:
Matches 20: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

```

```

Oy 1 ggaatcgtgtttgaagcag 22
    |||||111111111111111
Db 8488 GGAATCTGCTGTCGACGACAG 8509

```

```

RESULT 33
AE003686/c
LOCUS      AE003686      221888 bp      DNA      INV      05-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 11
            of 105, complete sequence.
ACCESSION  AE003686 AE002708
VERSION     AE003686.2 GI:10726424
KEYWORDS    HMG;
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 221888)
Adams,M.D., Ceiniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sulten,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazee,R.G., Champe,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Heitz,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Abdyan,A., An,H.J.,
Andres-Garcia,J., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Botchan,M.R., Bouck,J., Brockstein,P., Brothier,P., Butts,K.C.,
Busan,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delecher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Fertiera,S.,

```

```

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

```

```

Eienschmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jallali,M., Kalush,F., Kaipen,G.H., Ke,Z.,
Kennison,J.A., Kelchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Krautiz,S., Kulp,D., Lai,Z., Lasko,P., Lai,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milhina,N.V., Mobarry,C.,
Morris,D., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusken,D.R., Pacleb,J.M., Palazzolo,M., Peltan,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reiter,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Stimpson,M., Skupski,M.P., Smith,T., Spier,E., Spredling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 221888)
Adams,M.D., Ceiniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7299308.
Location/Qualifiers
1..221888
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
complement(join(232..838,1727..2384,3247..3489,3547..4303,
4652..4762,5777..5989,8720..8991))
/gene="best"
/product="CT19608"
/db_xref="FLYBASE:FBan0006264"
/db_xref="FLYBASE:FBgn0040238"
complement(<232..>8991)
/gene="best"
/map="85F13-85F15"
/db_xref="FLYBASE:FBan0006264"
/db_xref="FLYBASE:FBgn0040238"
complement(join(513..838,1727..2384,3247..3489,3547..4303,
4652..4762,5777..5847))
/gene="best"
/note="best gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0006264"
/db_xref="FLYBASE:FBgn0040238"
/protein_id="AAFS4503.1"
/db_xref="GI:7299309"
/translation="MTITTYGEVATTCGRCFCFLKLLRMGSIYKLVMDLPLITY
YAINWVYRGCLNPQKTEFAIVQICSYRELPLPLSVLGLFYISVWTRMNYSTIP
WPDPLAEVSSNVSGODEGRGMRTIMRYVCLDLMVLANVSPRYKRPGLNIVE
AGLNDKEKTIETMNRKAFPRPSKHMPIWASIIIRAKERIRDDFAVKTIIDEL
NKFPGGCLISYDTISVPLVYOVVLAIVSYFLFCGCGOQOQMGVNTYLVNKV
DLVFPVFTLQFFYMGMLKVAESLIIPREDDDPVMNMNDPNDVQVNTYLVNKV
DHPELIDQWDEVPNELPYTTLIAERFRENHEPRSTAKLEVRNMAWSTBSKVRD
EMADADGSIHFSANCKMRDSSSLVSVSGTISRVTVAALKRFLRSDRSRGSATP
PSQDQCPKFPASASASLSGAVSASVTSVSGTISRVTVAALKRFLRSDRSRGSATP
DPRNVNDIIFAQTSAGTSGPLQPPPAHSEPYDIDSPSYNRASQSYEPNIFPFGC
VDALISTSAPAGSGSPLLISNAATAPSPVSSKSLYDPQCGASRFVSSMDLRSTD
LIGDAVQPEDEGDDPDKLKAERKEKELMGOQNLARTSTAGMATATVPKVPMPV
NVAVQAOLOPVASSADLLAGDOFSSTMKKSFDAING"
complement(join(<10693..10854,11008..11150,11212..11403,
11465..11562,11780..11858,11924..12033,13138..13244,
13478..>13489))

```

mrna

CC	*	115132	143998:	contig of 28867 bp in length		
CC	*	143999	144098:	gap of 100 bp		
CC	*	144099	191736:	contig of 47638 bp in length.		
XX	Key	Location/Qualifiers				
PH	source	1. .191736				
FT		/chromosome="X"				
FT		/db_xref="taxon:9606"				
FT		/organism="Homo sapiens"				
FT		/map="X"				
FT		/clone="RP11-306E15"				
FT		/clone_lib="RP11-11 Human Male BAC"				
FT	misc_feature	1. .1537				
FT		/note="assembly_fragment"				
FT	misc_feature	1638. .3266				
FT		/note="assembly_fragment"				
FT	misc_feature	3367. .7315				
FT		/note="assembly_fragment"				
FT	misc_feature	7416. .12732				
FT		/note="assembly_fragment-clone_end:17-vector_side:right"				
FT	misc_feature	12833. .18750				
FT		/note="assembly_fragment"				
FT	misc_feature	18851. .24921				
FT		/note="assembly_fragment"				
FT	misc_feature	25022. .30115				
FT		/note="assembly_fragment"				
FT	misc_feature	30216. .38744				
FT		/note="assembly_fragment"				
FT	misc_feature	38845. .45042				
FT		/note="assembly_fragment"				
FT	misc_feature	45143. .53343				
FT		/note="assembly_fragment"				
FT	misc_feature	53444. .59314				
FT		/note="assembly_fragment"				
FT	misc_feature	59415. .66278				
FT		/note="assembly_fragment"				
FT	misc_feature	66379. .73615				
FT		/note="assembly_fragment"				
FT	misc_feature	73716. .83850				
FT		/note="assembly_fragment"				
FT	misc_feature	83951. .95698				
FT		/note="assembly_fragment"				
FT	misc_feature	95799. .115031				
FT		/note="assembly_fragment"				
FT	misc_feature	115132. .143998				
FT		/note="assembly_fragment-clone_end:SP6-vector_side:left"				
FT	misc_feature	144099. .191736				
FT		/note="assembly_fragment"				
XX	Sequence	191736 BP; 59272 A; 36008 C; 35590 G; 59163 T; 1703 other;				
Query Match						
Best Local Similarity		75.2%;	Score 18.8;	DB 33; Length 191736;		
Matches 20; Conservative 0;		Mismatches 2;	Indels 0;	Gaps 0		
OY	1	ggatcgctgtttgaagcgag 22				
Db	96915	GGATCTGCTGTCTGACGACAG 96936				
RESULT 32						
LOCUS	AL592563	193594 bp	DNA	HTG 13-JUL-2001		
DEFINITION	Homo sapiens chromosome X clone RP11-306E15 map q13.3-21.2, ***					
ACCESSION	AL592563	AC025411	SEQUENCING IN PROGRESS ***; 14 unordered pieces.			
VERSION	AL592563.2	GI:15029540	HTG: HTGS_PHASE1; HTGS_DRAFT.			
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					

```

REFERENCE      Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
AUTHORS       1 (bases 1 to 193594)
TITLE         McElay,K.
JOURNAL       Direct Submission
              Submitted (12-Jul-2001) Sanger Centre, Hinxton, Cambridgeshire
              CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
              requests: clonerequests@sanger.ac.uk
              On Jul 30, 2001 this sequence version replaced gi:17283744
              gi:14715681.
COMMENT       Draft Sequence, produced by Whitehead Institute/MIT Center for
              Genome Research, 320 Charles Street,
              Cambridge, MA 02141, USA
              http://www-seq.wi.mit.edu
              -----
              Genome Center
              Center: Sanger Centre
              Center code: SC
              Web Site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              -----
              Project Information
              Center project name: ba306E15
              -----
              Summary Statistics
              Assembly program: XGAP4; version 4.5
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Consensus quality: 186482 bases at least Q40
              Consensus quality: 189139 bases at least Q30
              Consensus quality: 190831 bases at least Q20
              Insert size: 192294; sum-of-contigs
              Insert size: 190691; 9.3% error; agarose-fp
              Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
              coverage: 4.58x in Q20 bases; agarose-fp
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 14 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              1       7271: contig of 7271 bp in length
              *
              7272 7371: gap of 100 bp
              *
              7372       26649: contig of 19278 bp in length
              *
              26550 26748: gap of 100 bp
              *
              26750       30691: contig of 3942 bp in length
              *
              30692 30791: gap of 100 bp
              *
              30792       59813: contig of 29022 bp in length
              *
              59814 59913: gap of 100 bp
              *
              59914       62315: contig of 2402 bp in length
              *
              62316 62415: gap of 100 bp
              *
              62416 78769: contig of 16354 bp in length
              *
              78770 78869: gap of 100 bp
              *
              78870       90625: contig of 11756 bp in length
              *
              90626 90725: gap of 100 bp
              *
              90726       96694: contig of 5969 bp in length
              *
              96695 96794: gap of 100 bp
              *
              96795       105609: contig of 8815 bp in length
              *
              105610 105709: gap of 100 bp
              *
              105710 115711: contig of 10002 bp in length
              *
              115712 115811: gap of 100 bp
              *
              115812       163536: contig of 47725 bp in length
              *
              163537 163636: gap of 100 bp
              *
              163637       171008: contig of 7372 bp in length
              *
              171009 171108: gap of 100 bp
              *
              171109       177629: contig of 6521 bp in length
              *
              177630 177728: gap of 100 bp
              *
              177730       193594: contig of 15865 bp in length.
              *
              Location/Qualifiers
              1..193594
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="X"
              /map="q13.3-21.2"

```

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source

1. 190673

Location/Qualifiers

/organism="Drosophila melanogaster"

/strain="y; cn bw; sp"

/db_xref="taxon:7227"

/chromosome="3R"

/map="86A-86A"

/clone="BACR03112 (D971)"

/clone.lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"

PBAC3.6)

BASE COUNT 54044 a 41124 c 41321 g 54184 t

ORIGIN

Query Match

Best Local Similarity 90.9%; Pred. No. 3.4e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gactcgtcgttgaagcagc 23

||||| ||||| ||||| |||||

DB 74835 GATCTGCTCTTGAGCCGACG 74814

RESULT 31

AC025411

ID AC025411

XX standard; DNA; HTG: 191736 BP.

XX AC025411;

XX AC025411.2

SV

DT 14-MAR-2000 (Rel. 63, Created)

DT 23-MAR-2000 (Rel. 63, Last updated, Version 2)

XX

DE Homo sapiens chromosome X clone RP11-306E15 map X, WORKING DRAFT SEQUENCE,

DE 18 unordered pieces.

XX

XX HTG: HTGS_DRAFT; HTGS_PHASE1.

XX

OS Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

XX

RN 11

RP 1-191736

RA Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,

RT "Homo sapiens chromosome X, clone RP11-306E15";

unpublished.

XX

RN 2

RP 1-191736

RA Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,

RA Anderson S., Baldwin J., Barina N., Bastien V., Beda F., Boguslavsky L.,

RA Boukhalter B., Brown A., Burkelt G., Campopiano A., Castile A., Choepel Y.,

RA Colangelo M., Collins S., Collymore A., Cooke P., Dearlano K., Dewar K.,

RA Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., Fitzhugh W.,

RA Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L.,

RA Grand-Pierre N., Grant G., Hagos B., Heatord A., Horton L., Howland J.C.,

RA Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., LaRoque K.,

RA Lamazares R., Landers T., Lehoczy J., Levine R., Liu C., Liu G.,

RA Locke K., MacDonald P., Marquis N., McCarthy M., McEwan P., McGurk A.,

RA Mlenka V., Morrow J., Murphy T., Maylor J., Norman C.H., O'Connor T.,

RA O'Donnell P., O'Neill D., Olivari T.M., Oliver J., Peterson K., Pierre N.,

RA Pisanic C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A.,

RA Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N.,

Stojanovic N., Subramanian A., Talamas J., Tesfaye S., Theodore J.,
Tirrell A., Travers M., Trigilio J., Vassiliev H., Viel R., Vo A.,
Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A.,
Zody M.;

Submitted (08-MAR-2000) to the EMBL/GenBank/DBJ databases.
Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA

On Mar 22, 2000 this sequence version replaced gi:7210134.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8131

Center clone name: 306_E_15

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 181491 bases at least Q40

Consensus quality: 186530 bases at least Q30

Consensus quality: 188547 bases at least Q20

Insert size: 192000; agarose-fp

Insert size: 190036; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 18 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1537: contig of 1537 bp in length

1538 1637: gap of 100 bp

1638 3266: contig of 1629 bp in length

3267 3366: gap of 100 bp

3367 7315: contig of 3949 bp in length

7316 7415: gap of 100 bp

7416 12732: contig of 5317 bp in length

12733 12832: gap of 100 bp

12833 18750: contig of 5918 bp in length

18751 18850: gap of 100 bp

18851 24921: contig of 6071 bp in length

24922 25021: gap of 100 bp

25022 30115: contig of 5094 bp in length

30116 30215: gap of 100 bp

30216 38744: contig of 8529 bp in length

38745 38844: gap of 100 bp

38845 45042: contig of 6198 bp in length

45043 45142: gap of 100 bp

45143 53343: contig of 8201 bp in length

53344 53443: gap of 100 bp

53444 59314: contig of 5871 bp in length

59315 59414: gap of 100 bp

59415 66278: contig of 6864 bp in length

66279 66378: gap of 100 bp

66379 73615: contig of 7237 bp in length

73616 73715: gap of 100 bp

73716 83850: contig of 10135 bp in length

83851 83950: gap of 100 bp

83951 95698: contig of 11748 bp in length

95699 95798: gap of 100 bp

95799 115031: contig of 19233 bp in length

115032 115131: gap of 100 bp


```

VERSION      AF052042.1  GI:3916123
KEYWORDS
SOURCE
ORGANISM     Norway rat.
              Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE    1 (bases 1 to 2472)
AUTHORS     Dovatz, S., Gilbert, K.A., Petrovic-Dovatz, L. and Rannels, D.E.
TITLE       Isolation, cloning and characterization of a novel rat lung zinc
JOURNAL     Biochim. Biophys. Acta 1442 (2-3), 380-388 (1998)
MEDLINE     99023763
REFERENCE    2 (bases 1 to 2472)
AUTHORS     Dovatz, S., Gilbert, K.A., Petrovic-Dovatz, L. and Rannels, D.E.
TITLE       Direct Submission
JOURNAL     Submitted (03-MAR-1998) Cellular and Molecular Physiology H-166,
              Penn State University, College of Medicine, PO Box 850, Hershey, PA
              17033, USA

FEATURES
source       1..2472
              /organism="Rattus norvegicus"
              /strain="Sprague-Dawley"
              /db_xref="taxon:10116"
              /tissue_type="Lung"
              1..2472
              /gene="RLZF-Y"
              171..1781
              /gene="RLZF-Y"
              /codon_start=1
              /product="zinc finger protein Y1"
              /protein_id="AAC78780.1"
              /translation="MAGSGVAPPPLGVGLCAVAVKVEDPSGQSPGSGGDMQENPSR
              KYFQOLYQVAVGPEEALSRMLCEKRWMLPELRSKQIMELVLVDFLITLRELOA
              RHPDCEGSEGAALARTLORALDGLSLOSPATFKDYASLITMEBQILAAKGFQ
              RESIKDPGSGVGLTETKATVTYVILKQESKABESQAWLQEVYSGKVPVTKGDTW
              EDMEERLPKAEELPLQSSPEEGKRTAIPLHLSVSKDESQKDNFENSGSLVQGH
              QTAEGLVNCEGCEDEHKGGLHACGVYKPHSSVDNALGLIESQHNFEQEGPYCDNCE
              KRRFORSDLRKHQRTHTGKPYQCEGCKFSQSAALVYKHQRTHTGKPYACPCGEC
              FRRSSLSRHQRTHTGSEKPYCKCEGCFPHISLSLKHQRLHKGSRPHCKEVCCKSPKQ
              RSDLFKHQRTHTGSEKPYCKEPCGRRFSQSAATLKHQRTHTGKPYCKPCQCGRRFROST
              HLYRKHQRTHTGSEKPYCKEPCGRRFSQSAATLKHQRTHTGKPYCKPCQCGRRFROST
              309..566
              /gene="RLZF-Y"
              /note="Region: leucine-rich domain"
              576..689
              /gene="RLZF-Y"
              /note="Region: KRAB-A domain"
              1158..1709
              /gene="RLZF-Y"
              /note="Region: C2H2 zinc finger"

BASE COUNT  665 a 609 c 689 g 509 t
ORIGIN
Query Match      75.2%; Score 18.8; DB 10; Length 2472;
Best Local Similarity 90.98; Pired. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 gatcgtcttgaagcagc 23
      ||||| |||| |||||
Db 419 GATCTGCTTTTGAGCGCAGC 398

RESULT 28
LOCUS       AF087452      8371 bp      DNA      BCT      23-SEP-1998
DEFINITION  Bacillus subtilis peptide synthetase (fenc) gene, complete cds.
ACCESSION   AF087452
VERSION     AF087452.1  GI:3643186
KEYWORDS
SOURCE      Bacillus subtilis.

```

```

ORGANISM     Bacillus subtilis
              Bacteria; Firmicutes; Bacillus/Clostridium group;
              Bacillus/Staphylococcus group; Bacillus.
REFERENCE    1 (bases 1 to 8371)
AUTHORS     Lin, T. and Liu, S.
TITLE       Molecular Characterization of peptide synthetase fenc gene in
              Bacillus subtilis F29-3
JOURNAL     Unpublished (1999)
MEDLINE
REFERENCE    2 (bases 1 to 8371)
AUTHORS     Liu, S.
TITLE       Direct Submission
JOURNAL     Submitted (26-AUG-1998) Shih-Tung Liu, Microbiology and Immunology,
              University of Chang, Gung, 259 Wen-Hwa 1st Road, Kwei-San,
              Tao-tuan, Taiwan, 333, R.O.C

FEATURES
source       1..8371
              /organism="Bacillus subtilis"
              /strain="F29-3"
              /db_xref="taxon:1423"
              /clone_id="pf6A5"
              692..8371
              /gene="fenc"
              692..8371
              /gene="fenc"
              /codon_start=1
              /transl_table=11
              /product="peptide synthetase"
              /protein_id="AAC36721.1"
              /db_xref="GI:3643187"
              /translation="MKENTYGLTHAORRWTFELLEPGTSTICNTLACVKGPGVDLLA
              LQALNLSISRNDALRPFOLTEGSESPRLYAKSLSEITDFTAENETREBOWI
              HOARLPKLDSPKXOPYLLRHPHEHMLAKFHHTIMDSLSLGNQITIDLYOKI
              KKGEPVDEQSESTYSTEKEQYLOSSREFEDHLEWQYQOQPVVHSLGQSTQSK
              QSTASRDTITLSPYLAHTQACEQKISMISLEMASPYICISNTSKTLAIGTY
              GRRSGKAEKDMGMFVSSLPILKTYDSADPLSEFARSYGRQLSYMRQRPYMLVN
              ELRKHKYLMNIGISMQYHPLOMKKADLVNLTALYSGYTNLSVQIDERDTGA
              IOLNFDYQNTLESLDIKRIQHHLLTLENLONPHSTIKELDMAEKREKRYXEN
              RKAVSPYSTHGLFEQOALTPRTALRPSGSGTYAOQDMYANRLARTANGITK
              ESIQVLSERSPDMTAYLAVLKAGATVPIDVYPERLSYMLKDSQAAALLPQGL
              KAPVSGCTLEVDMTSLANESYNSHQADSDSLAVLYTSGSTQGRGVAYAHVLA
              VSELDGMQRPLEEDDIIMKTSRSPASVQWLFMWLLSASAYLLPSGWEKDPALM
              KAIREEGVYTAHPIPMINSTRFLDQAEIAPRSILKRVAPAGEPLAPHTAAFSALBE
              TSAVGYGPTATVDAFAFYACDPELDRKRLPIGKPYGARGVILVDPHVAOPGVA
              CELYTAGAVAGVYLNRPETLEERLEPRLPGKPYKPYGDLAHMLDGOVERIGRID
              DOVKIRGRIEPEGCEALRSIECYRDAAYVYRSESGAEICATAGELNEMKLSQLE
              TLLRPYMPAHITIELQMPVTPSGKLDKRALPADGADRETYAPRNLNEMKLSQLE
              BEVLSGFGVGHNDNFDRGNSLKATALVSLAEFGQVPLKQVFAPHTVEGLASVY
              SEGTSPYEALIKPAKCEPTYVSSQKRMVYLQLEDGCTGYNMPAVLEEGKIDLR
              MEAVKRLIKRHESLRTSETYDIDGEPYQRLHDEVPEPTFPSTVGLGGTEQDEAAAFIO
              PFELSQADPLFRSGVYKVSIDGRHLLLVDMHNLISDGVSNYLLIREGELVANOELRALH
              IOYKDYAVMRGEGPKGDTYKTOFAYWYIKOLEGELPYLDIPADYARPPYRSPAGDOVSF
              MUDQELTAGLHKIARENGSTLYMVLAAATYATLARLSQEDDIIVGSPPLAGPRHDEP
              ILGMFVNTLALRTPEGRKRPPTQFLQEVRETALEYBHDYPEBELVDKLGVRDMKR
              NPEVDAMPILQNMKODIHLDIKARPNVTHQISVDEDMTLMASDGVIKDMEFST
              DVFMRKTLTERTFAHTEFLRAATSNPNTTISQVDSLSEKEKQKILILNTHVESCOT
              DTVFRMFKEARETEPEHIAVIDEKOISYHLEKRNRLARTLOEGKEPQIPVATL
              AERSIDALVGLIAWAKAGVYIPIDSHYPARAVYILKDSGAEILLIIONVYKLIJAS
              DICDSIRPCIDDESYPSSKMHLSPPAPDESAVITYTSGTGAQPKVITYTHRRFANA
              VIAMRYIOLDMOYRILOMASFSFDPVPSGLARTLANGVLYICDPETLEPRLA
              LMRNQRITMEATPALVFMELVYRNQSLIPDLDILILSDSKAKADFTKLDRECN
              KMLNLSYGVTEATLIDSSYEMNGEYSGQSVPIGIPLPNVCIVLSQDOLOIPGI
              AGELCIAGAGVAKGYHGRFELTEKKEPTENFVPEGRILRGDLACVMPNCTLRIGRI
              DHQVAINQYRIETEEIESVLQDTGLNVAVAVONONGOARLAAYILPDPADPTTLAR
              GALSMLPRAVWPAHIIPLFETMPLTLNGKIDRLAPARNVARSPTAPYHEHLSQMA
              OIMEDVLSVRSVGHIDSPFELGDSIKALQVSRALAEFGSMYISDLPAPYHEHLSQ
              LKLTTEHDAVRAWFTQDHTGRVQVNRGQITENELFGHLIDWKEQAKRLNLEK
              PAEEIYVDSNMNVEBQPLQAGLEKTRKEDHLLITLHNAVGVSGNRILLEDLAAY
              QOSLEQTEIKLPKSDSYLYANGLAIOASRQLSLETITWQTLDAHNFELPKDAE
              SDRHQSAAETAAEVLASDWTKLLEFQOAGYGDANLTLTALGMLAMKESGHEQIYI
              SAEHGCVGHVNPIDISRTVGMFSTIYPIILLMSAPSSSDQIDGARYIKCKKDLRRYP
              HKGTGYGLVQKGAIKKPEPEISPNYIGQLSEBSSSLPOLSSYQPTREYTAGREKEY
              ELVINALITDGRLOYKAVYTRVRENTIQCLMDSFHSLLIETIIDCTKKEREKTLSD

```

CDS

complement(<3698..>. >4543)

/gene="CG14065"

/map="98C3-98C3"

/db_xref="FLYBASE:FBan0014065"

/db_xref="FLYBASE:FBgn0039579"

/evidence=not_experimental

complement(3698..4543)

/gene="CG14065"

/note="CG14065 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0014065"

/db_xref="FLYBASE:FBgn0039579"

/evidence=not_experimental

/protein_id="AAF56784.1"

/db_xref="GI:7301663"

/translation="MSSKEETPTGCAAPVATVAAGSGA""TSYANYVNILERKKSSTS
NTTNTTAATYADNDENPNLKSEKSMSEFAAAEAEGESITVGSVSLAQIEGRRAAS
GGDNTAEKMSSTLDNNDFPVYLSHHRRDKRKAREPDDOPAGCGGDGPQOAGR
RVPGSDTERKSVRPARRGSPPKSKAAGAAGSNAAGVSSAVGCPGVHKHRDPT
SPSLAIEDAGSSGNENAKSADDDQAAGEKGAAGAAPPKRFTIADPPRVNMAKVSIDMT
EISLSD"

CDS

complement(join(7319..7693,7896..>9739))

/gene="CG1345"

/product="CT3038"

/db_xref="FLYBASE:FBan0001345"

/db_xref="FLYBASE:FBgn0039580"

/evidence=not_experimental

complement(<7319..>9739)

/gene="CG1345"

/map="98C3-98C3"

/db_xref="FLYBASE:FBan0001345"

/db_xref="FLYBASE:FBgn0039580"

/evidence=not_experimental

complement(join(7319..7693,7896..9572))

/gene="CG1345"

/note="CG1345 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0001345"

/db_xref="FLYBASE:FBgn0039580"

/evidence=not_experimental

/protein_id="AAF56785.1"

/db_xref="GI:7301670"

/translation="MCGIFAYLVLTPEKSRQEVLDLLIQGLRELYRGYSTGIAIDA
INSGEOSIMLYKPRGKVLEDAVAECRODPSLPIDHIIAGTHRMATHGPSV
NSHPQRSDNPFPVVNGCATITNKDYDTTLKEKGVESTDPTEVIATKIIVNHMOOH
PGTFRELIEDQALQQELACFALAFKRSKFPEPCVASRSGSLVIGTKAKIKIADHIP
LIYAARHHQGPQOOATQLDPGCCNEFGPLEKEVEITFMKGTQIMGANDVFMLKFIEO
LEDDEAVAANKSDGLTSHRLKNSJSDPARLETILKMETQIMGANDVFMLKFIEO
PESSVNTVMGMRRPDTQTVVLQGIKEYIPETKRCRIMLIACGTSYSNAVTROLLEE
LFPLPWNLIASDELDRNTPTRFDVCVSAATSOFSLMFALVMSDEBLISQURLEI
TVGSSIICRSCHSGVHNAGPEIGVASTAFYSOTLSMGFALVMSDEBLISQURLEI
IQLSOLDIRFTVKLINSONVOOLAKEILEHRSLLIMGREFNATCLEGALKVELLEY
MSEGLFLASELIKPLALVDKMPRLMYLHVDPYVTKCMNALDOOVTSKSGPIILICE
GDNEMTSFTRLQIPRTVDCIQGVLTVPIDQLSYHIALVLRGCDCVDCPRMLAKSVTY
E"

mRNA

join(10379..10764,11398..11485,11653..11701)

/gene="CG1866"

/product="CT37544"

/db_xref="FLYBASE:FBan0001866"

/db_xref="FLYBASE:FBgn0039581"

<10379..>16982

/gene="CG1866"

/map="98C3-98C4"

/db_xref="FLYBASE:FBan0001866"

/db_xref="FLYBASE:FBgn0039581"

join(10387..10764,11398..11485,12211..12240,12306..12445,
12505..12652,12709..12835,12906..13100,13161..15196,
16982..16982)

/gene="CG1866"

/product="CT5760"

/db_xref="FLYBASE:FBan0001866"

/db_xref="FLYBASE:FBgn0039581"

join(10461..10764,11398..11485,12211..12240,12306..12445,

CDS

	12505 . 12652, 12709 . 12835, 12906 . 13100, 13161 . 15041)
gene	/gene="CG1866"
note	/note="CG1866 gene product [alt 2]"
codon_start	/codon_start=1
db_xref	/db_xref="FLYBASE:FBan0001866"
db_xref	/db_xref="FLYBASE:FBgn0039581"
protein_id	/protein_id="AAFS6786.1"
db_xref	/db_xref="GI:7301671"
translation	/translation="MTVNRKRDAGATRRPCFEDISLGGLGMRITVELFENDVAPXTAEN FALCTGEGEGELITGKRIQYKCVI FHHVYKDFEYVADGFSNGNGGSEISYGFED ESFEKHKDRPFLILSMANRGKNTGSOFFITTPPAFHLNTHIVYFGVATISGDELIVOLE GLPVNRNSRPLQDAAIANCGELVROTKAKKKRKRSTATEDNSSESEALAKVYRK AKKKRSKRDTSQSOSDENETRSRGKRIHDQTPDQDDEDEREGELHPLVLTITIDNE IDEVNSKFLPRAERSKSIDLREERGAAGUORDGKQDQDRSRNNNGGWSKQAPPSRS GTVLGVGVYFRFPTPSRSRSTPTPPMHHAKRITKLSDLERIEESKREEMKKR EHEHRKREHETAKNPKOSFEFLSHASTYCGKITPDESPENHAKSKMDYDRDKAKRE KAGEKDARDYTPVYRKSMQNDNALDI EOOTESPAREDEEOLKYVPSKODIRVPS TSDRAKAKDERKPTTEKAKPKRSRSRSPROSAPARODSPARRRGQONONONO NONONONOLNCRNRNRPADRDNRRSKSDNEMETRRYPLSTIKGSPSCDRRQOR SSSONRQDRPSQOROLISGPKRKQDSKSKRPNSTETKWKDSEPKRQDSHDKRRAS PWNKRDSPPARKRKDSANRRKREDSDQKIQDSRKQDSAGKKREDSKRRKQDSQK QOSPVRKSGSPGKRRQDSPEERKRDPDRRRDPSAPRRQESGRRQDSQDRRRES GNNRRSGHSMGHHSSRSRSPSTSRSRRRNSPLKLSALPMSERREKARREKQK RAVAAMLMKEHMRNPTAKPEEREMKQDDLDKEPTRELEINLEIKLAEITRLKLEF DSYDPPAAHSAGSASAKTADTSTGCAKSKROSSNNHREKKRHKKKSKSGSNDN"
join	join(10461..10764,11398..11485,11653..11695)
gene	/gene="CG1866"
note	/note="CG1866 gene product [alt 1]"
codon_start	/codon_start=1
db_xref	/db_xref="FLYBASE:FBan0001866"
db_xref	/db_xref="FLYBASE:FBgn0039581"
protein_id	/protein_id="AAFS6787.1"
db_xref	/db_xref="GI:7301672"
translation	/translation="MTVNRKRDAGATRRPCFEDISLGGLGMRITVELFENDVAKTAEN FRLCLTGEKGFGLITGKRIQYKCVI FHHVYKDFEYVADGFSAGNGGSEISYGFED ESEFEKHKDRPFLILSMANRGKNTGSOFFIISPALMCELAHSH"
join	join(<22268..22577,22633..23528,23596..23703)
gene	/gene="Or98b"
product	/product="Cm5076"
db_xref	/db_xref="FLYBASE:FBan0001867"
db_xref	/db_xref="FLYBASE:FBgn0039582"
<22268..>	<22268..>23703
gene	/gene="Or98b"
note	/note="CG1867"
map	/map="98C3-98C4"
db_xref	/db_xref="FLYBASE:FBan0001867"
db_xref	/db_xref="FLYBASE:FBgn0039582"
join	join(22268..22577,22633..23528)
gene	/gene="Or98b"
note	/note="Or98b gene product"
codon_start	/codon_start=1
db_xref	/db_xref="FLYBASE:FBan0001867"
db_xref	/db_xref="FLYBASE:FBgn0039582"
protein_id	/protein_id="AAFS6788.1"
db_xref	/db_xref="GI:7301673"
translation	/translation="MLTDKFLRLRQSAFLRLGLLELHEDVDYGHYRPRSTICILTSVAS FMPLTIAFGLONVQNEOILDTSLCSVLVDLALCKICLDFMLTKDFKFLILGQFVCVLQ
Query Match	76.8%: Score 19.2: DB 3: Length 239171:
Best Local Similarity	87.5%: Pred. No. 2.2e+02:
Matches	21: Conservative 0: Mismatches 3: Indels 0: Gaps 0:
QY	2 gatctgctctttgaagcgacgacg 25 DB 164426 GATGTCCTGTTTGAAGCTCAGCAG 164449
RESULT 27	
LOCUS	AF052042 2472 bp mRNA ROD 23-NOV-1998
DEFINITION	Rattus norvegicus zinc finger protein Y1 (RNLZ-F.Y) mRNA, complete cds.
ACCESSION	AF052042

DEFINITION Drosophila melanogaster, chromosome 3R, region 98C-98D, BAC clone BACR14L13, complete sequence.
 ACCESSION AC008028
 VERSION AC008028.3 GI:12957647
 KEYWORDS HTG
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster

REFERENCE Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta: Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
 1 (bases 1 to 197348)
 AUTHORS Celinker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, R.H.,

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brannon, R.C., Rogers, Y., An, H., Baldwin, D., Bazon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Chape, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorset, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D., Fierres, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzales, M., Hock, J., Hoskins, R.A., Hostin, D., Howland, T.J., McInosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phoonanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 Sequencing of Drosophila chromosome 3R, region 98C-98D
 Unpublished
 2 (bases 1 to 197348)
 TITLE Celinker, S.E., Abdayani, A., Arcaina, T.T., Baxter, E., Blazet, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummst, S.R., Kaitra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svitskas, R.R., Wan, R.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
 Direct Submission
 Submitted (12-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
 COMMENT On Feb 17, 2001 this sequence version replaced gi:5670462.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

FEATURES
 source
 1. 197348
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /strain="y: cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="98C-98D"
 /clone="BACR14L13 (D789)"
 /clone_lib="RPCT-98 (Roswell) Park Cancer Institute
 Drosophila melanogaster BAC library, partial ECORI in
 PBAC3.6)"
 BASE COUNT 54902 a 42954 c 44141 g 55351 t
 ORIGIN

Query Match 76.8% Score 19.2; DB 3; Length 197348;
 Best Local Similarity 87.5% Prid. No. 2.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 2 gatctgctgttgaagcagcagc 25
 ||| ||||| ||||| |||||
 Db 188662 GATGTCCTGTTGACGACGAC 188685

RESULT 26
 AE003765
 LOCUS
 DEFINITION Drosophila melanogaster genomic scaffold 14200013366035 section 90 of 105, complete sequence.
 ACCESSION AE003765 AE002708
 VERSION AE003765.2 GI:10726824
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster

REFERENCE Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta: Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
 1 (bases 1 to 239171)
 AUTHORS Adams, M.D., Celinker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brannon, R.C., Rogers, Y.H., Blazet, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, M.K., G.L., Abril, J.F., Abdayani, A., An, H.J., Andrews, P., Fink, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottler, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrelle, A., Fleischmann, W., Foster, C., Gabrielian, A.E., Gary, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Hock, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalish, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kuip, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Matel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacle, J.M., Palazolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Sanders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spindling, A.C., Stapleton, M., Strong, R., Sun, E., Svitskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, O.A., Ye, J., yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of Drosophila melanogaster
 Science 287 (5461), 2185-2195 (2000)
 2 (bases 1 to 239171)
 TITLE JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7301668.
 Location/Qualifiers
 1. 239171
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="3R"
 complement(13698..>4543)
 /gene="CG14065"
 /product="CT33634"
 /db_xref="FLYBASE:FBgn0014065"
 /db_xref="FLYBASE:FBgn0039579"
 /evidence=not_experimental

mRNA


```

PF      24-DEC-1984 JP 1984272605
PI      HIROMO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12P13/18, PC
C12P13/19);
CC      strandedness: Double;
CC      topology: Linear;
CC      hypothetical: No;
CC      anti-sense: No;
CC      *source: strain=Escherichia coli Ns101;
CC      feature is identified by experimental;
FH      key
FH      Location/Qualifiers
FT      CDS
FT      42..1079
FT      Location/Qualifiers
FT      1..1121
FT      /organism="Escherichia coli"
FT      /db_xref="taxon:562"
BASE COUNT      289 a 296 c 325 g 211 t
ORIGIN

Query Match      93.6%; Score 23.4; DB 6; Length 1121;
Best Local Similarity 96.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 ggatcgtcgttgaagcagcagcag 25
        |||||||
Db      104 GGATCTGCTGTTTGACGCCACGACG 128

RESULT 17
AF248314
LOCUS      8227 bp DNA BCT 24-JAN-2001
DEFINITION      Uncultured bacterium pCosAS1 urocanase-like protein (hutu) gene,
partial cds; histidine ammonia-lyase-like protein (hutu) gene,
amino transferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
synthetase (bioF), biotin biosynthesis BioC-like protein (bioC),
and dehydrobiotin synthase BioD (bioD) genes, complete cds; ABC
transporter-like protein (elisa) gene, partial cds; and unknown
gene.
ACCESSION      AF248314
VERSION      GI:12407610
KEYWORDS
SOURCE      uncultured bacterium pCosAS1.
ORGANISM      Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 8227)
AUTHORS      Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streitt, W.R.
TITLE      Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
JOURNAL      Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
PUBMED      11134332
2 (bases 1 to 8227)
AUTHORS      Entcheva, P., Liebl, W. and Streitt, W.R.
TITLE      Direct Submission
JOURNAL      Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
FEATURES
source
1..8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/notes="unknown organism, cosmid clone derived from
environmental consortium"
<1..502
/gene="hutu"
<1..502
/gene="hutu"
/notes="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"

gene
CDS
/protein_id="AAG53593.1"
/db_xref="GI:12407618"
/translation="PRERISFGQPARICWVQLEWRQKIGLAFNEMVRGEVSAPVI
GRDHLDSGSVASPNRETEAMODGSDAVSDMPLNALLNTAGATVWSLHAGGVMGF
SOHAGVAVICDGTDDAARLARVLHNDPATGVMRRADAGYDIADCAKEGGLNLMIA
ATOGOH"
502..2031
/gene="hutu"
502..2031
/gene="hutu"
/notes="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAG53586.1"
/db_xref="GI:12407611"
/translation="MNAULTFGTLTQLAQWMOOPLQLTDESANEAINDSVACVE
AIVAGRTAVGINTGGLAOTRIATHTDLENLORSIVLSHAGVGEPLDDIYRLMAY
LKINSLARGFSGIRLSVTOALITAVNAGVSVDPKSGSVGSLAPLAKHSLTLTLEE
AVYGCALTTAEVLGSRPFDARIEHVRGQKIDAALEFRHVLDTDSALASHNCKK
VODPYSLRCQPGVWGACUOMQVAEVLIVSNVSDNPVLFAAENEVEGRNFHAP
VAMADNLALAIAGLSEERRIALMDKHSQILPPLVLRGVNSGMIAOVYAAAL
ASENKGILCHPHTSVDKPCRSANOEDHVSAPAPRRLPMAGNTRGVLAWEALAC
OCADLRDGTSPLELQARQSGEGVYAHYDDDRFPAPDIEAIALSLKGLSVGLLPAP
1"
complement(2130..2606)
/notes="ORF1: similar to hypothetical 17.1 kDa protein in
modc-bioA intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG53587.1"
/db_xref="GI:12407612"
/translation="MKLISQDLRGDKLPLKRVNMGVEGDNISPLHAADEVPSGTR
SEVYVQYDPDAPDPSGMMHWIVANIADTRVLPGSGSSSLVALPEGAVERTDFGKAG
YGCAPPKGTHRYIFVHALVERIEVEDEBSAGMVFNVHFTLSASITMYS"
complement(2651..3940)
/gene="bioA"
complement(2651..3940)
/gene="bioA"
/notes="7,8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase BioA"
/protein_id="AAG53588.1"
/db_xref="GI:12407613"
/translation="MTTDDLAFLDRQIHMPYTSMTOPLPVVEAHGCELTLAGEO
LVDSMSWMAAIGHVNHPRILNAAMKAQIDRKSHYMFGGITHOPAVDLCRLVAMTPEE
LECVFPADSGVAEVAEMKMALOYMHAKGESROFIFFRNGYHGDFTGASVCDPDS
MHSLSMGVLPENLPAFAPOSREPGAMDEMVGIAPIAAMHREIAVLLEPIVOCAG
GMRVHPMLKRIKRMCDREGILLIAELIATGFCRKLIFACEAGCTPIITLIGAL
TGGMVMSATLTTRVOVAETISNGCAFGPMGPTFMCPRLACAVASESLALIESGEMD
OVAAIESLORELVARDSEFVADVRVLCAIGVETHPVNMALOREFVARVWVRP
FGKLITLMPYVIRPDLSIKLTVRAVDAVAPAHFTR"
4029..5069
/gene="bioB"
4029..5069
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG53589.1"
/db_xref="GI:12407614"
/translation="MAHDAWTLISOVATLPEKPLLELFEAQOIHQNHEDPOIOYST
LISKTCACPEDCYKQOSARRYTGLSEELMEVEEVLDESARQAKNAGSTRFCWPAW
KNHERMDPYLEQWVGVKAMGLQACTTGLTIDDDHOARLASAGLYVHNHNDPTPEF
VGNLITRTYQERLDITDKVADAGIKVCSGIGLETVDRAGLILQIQLNLPPTPES
VPINMLVKGVTPLADNDVYXAPDXIFTXSVAAVMKPTXIVRLSTGRQMNBDTOAMC
FMAGANSIFGCKLITTPNPEEDKDVQLFKKLILNQQRAVILGDNEDQHQLEQLFN
ADTDPFYNAATV"
5066..6223
/gene="bioF"

```

```

gene
1245..2114
/gene="ECS0754"
CDS
1245..2114
/feature="ECS0754"
/feature="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/transl_table=11
/translation="MSILIDKNTKVIICGFTGSGTSEHSEQAIVGTRMGVYPGK
GTHHGLPEVNTREAVATGATATVIVPAPPCDSILLEADIGIKLITITGIPIT
LDMITVKYKIDEPVARMICPNCPGIVTTPGECIKIGTPGHKPKGVIVSRGTLTVE
AKQTTDYGFGSTCVGIGCDPIPSNFTIDLEMPKDPQREAIYVMIGETGSAEEEA
AAVKEHVTRKPVGYIAGVTAPKGRMGHAGAITAGGKTADERFALEAGCVTVRS
LADIGELKTVLK"
2781..3686
/gene="ECS0755"
2781..3686
/gene="ECS0755"
/feature="ECS0755"
/feature="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g14170431sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/transl_table=11
/translation="MRGKIPKTELLVTEEVVAHESYTPAAEELATGOSAVFQVSAL
EFLRPLPFSKSKRKLPLNDGKYVIGYKTEFLNKLREPTNTITMOTPVQVILEAVN
PESTWMLIPNLAHEPKLHPDIIIVNHSIANGDPELVDAVIMRENCAPMAVEY
LDEEETLPYCSSLAMSDQKLSVAELLTELPDLOSRTITGMEEMFLSCVSSPLVN
NCPREDLSMLIAAVNSNLGVALDLPFAIHDLDSDGVIPCDVPIRTGNRIITWQE
EKSDSPHQFREFMLAKSVPOEM"
3720..4322
/complement(3720..4322)
/gene="ECS0756"
3720..4322
/complement(3720..4322)
/gene="ECS0756"
/feature="ECS0756"
/feature="note="probable cob(1)alamin adenosyltransferase, similar
to cob(1)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481sp|P13040|BRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/transl_table=11
/translation="MEARISTERRHROOKLKEOVDTRYAALTEKKGILLVFTGNGK;
KSTIAFGITVTRAGHGTGVAYOTIKGQDNSEYINLQPLGEPFHMGTPWETQNK;
QADIDAKKEVSESRMLADKRYDVLVDLEETMYALHYLDVEVYASLQNKPAQOSV
IVTGRCHSGLIMADTVSEIRPVKHAFDNGIOADPGIDW"
4332..5984
/complement(4332..5984)
/gene="ECS0757"
4332..5984
/complement(4332..5984)
/gene="ECS0757"
/feature="ECS0757"
/feature="note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11205981sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/transl_table=11
/translation="MSKPIIWOEPQLQNKDGEYITISDQHTITVTELDGEYIKTAPE
ALTRLARQAFYEASFPLRSALHDOVASILINDQASSNDKYVALQLLRLAEVSAKGLP
NCODTGTATIVASKGOITWGNDAEALSKGITYSFOENANLESONAPIDMTYEVNTQ
TNLPAOIDISAVAGDEYHFLCVKGGGSANKAAIKYOEKSLQPKRLTAFLIEKKSL
GTACPCPGYHAFVYVGLSDQTLKIKLASTYVYNLPTSGNDOGAFFDIELEKVL
EASQOPYIGAQREGGKIFANDIRYIRLPRGGSCPTIAMSASADSNITAKIKKHQITL
DIHAHAKIKARLSGEMPEYLRKHITVYAGPACTEPENMACGSLGPTTGGRMGVYIDTP
OAGGSLVIMLSKGRNSQVYDACGKHGGFNLSIGGAALLAOFYVKSILRCLYBELG
MEAVMMMEVENLAFILVDOKGNPFSEFEOOHRASCAPGH"
6092..7372
/complement(6092..7372)
/gene="ECS0758"
6092..7372
/complement(6092..7372)
/gene="ECS0758"
/feature="ECS0758"
/feature="note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g111214671sp|P24943|GLT_BACSF percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/transl_table=11
/translation="MKKISTITMILLALVYGMITGVVINTASPTAKLYAOEISTFT
TIFURLIKMIIAPLVSTLVAGTAKKMDAKLAGRTSKTIFPICSLSLAIGLITV
NFMPPGTGIVNEVMAHGAETGVVAEPTLKVFSHAFPTVDMAHNELLOLVSTI
FLGCSLTAIGESAIYHALDSLAHAMKLTGYVMFAPLTVFAALISALAEGLAMV
VSAGIIFMGFEFFETMLLVLLDGLAIYVGPICRLRLRLSEBALLAFTTSSSEAPF
GLEKLEPGVSPYIASEFVPIGSEFVLSGMAVCEFAVFIAQCNHILSIEQITM
LITMLTISKMGAGPRASVMVIAATNOFNIPAGJILMGVPPFLDMGSRASVMSN
AMGAAWYSRNDEGHEHRCGCKALKRPESVALP"
7533..7850
/complement(7533..7850)
/gene="ECS0759"
7533..7850
/complement(7533..7850)
/gene="ECS0759"
/feature="ECS0759"
/feature="note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

```

Query Match 100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatctgctgttgaagcagcag 25
Db 92373 GCATCTGCTGTTTGAACGACGACAG 92397

```

```

RESULT 16
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono, Y., Kojima, T. and Kimura, H.
TITLE DUPLICATION OF BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
PRODUCTION OF BIOTIN
PATENT: JP 1986149091-A 1 07-JUL-1986;
JOURNAL NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
COMMENT 07-JUL-1986

```

```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSMNSLTNRNLTPDNRLSSVRDAVAHSDNGAOV
KGNRTYRVAVATDNKFCVTRSHSGCFNLHLRLGMPKGLISRKIEVNLNAPVSAM
ERGLVHNSRDPDLPVADPPELSPVYDNRSLPGCVTKGCGNNAVYEDAPATYVLAK
FTTSGNSEVTESEVRCFNOYAGASAEKTYGNNNGDILICIRNDK INGESLTLSPAO
AEHATYDMPDLLECKGILFVDTTEINVLIDRAKNEFNFDISYNSVDSRWSSEQIMQ
SYHGKKDPLISVLSKTI"
/gene="20990"
/length=7160
/size=6462..7160
/gene="20990"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPPLSLVLYHQSRRSSMSDIDIESLVKASSVQMIKNPOLRF
QGDHNITVQOIEALDKIGSTETGRVLLNIESIRLSETVVHLNLSRLGVAHHD
IDAHNRGTGSDIFHCNINAVYPCGEGISVDDFATVFEHLHFNHNGRLKVS
SRASOKYSPILLEARTVGLGARSEVLESENKTHEIDGMPRTSYXPDSALLHDDNT
VSLGFOVRLHPIL"
/complement(7668..8144)
/gene="ybhB"
/note="20992"
/complement(7668..8144)
/gene="ybhB"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKRLSNDLKDQKLPRIHVPNGMGYDGNIIISPHLAMDVPACTK
SPVYCTDPAPATPGSGMMHVVNVLPAETRVLPQFGSGGLVAMPDGLQTRDFGKTG
YDGAAPPKGETHRYIFLVALDVERIDVDEGASGAMGVFVHFHSLASSTAMFS"
/complement(8203..9492)
/gene="b10A"
/note="20993"
/complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggacltgcgtttgaagcagcag 25
Db 9641 GGATCTGCTGTTTCAAGCGACGAG 9665

RESULT 15
AP002553
LOCUS 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553
VERSION AP002553.1
KEYWORDS GI:13360211
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:R1MD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.
REFERENCE 1 (sites)

```

AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shingawa, H.
Complete nucleotide sequence of the prophage ϕ 72-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Hayashi, T., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shingawa, H. and
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic *Escherichia coli* O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shingawa, H.
Complete nucleotide sequence of the prophage ϕ 71-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli*
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shingawa, H.
Complete genome sequence of enterohemorrhagic *Escherichia coli*
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shingawa, H. and
Hayashi, T.
Direct Submision
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.

COMMENT
FEATURES
source

Location/Qualifiers
1..297816

CDS

/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="R1MD 0509952"
/db_xref="taxon:8334"
79..1245
/gene="ECS0753"
79..1245
/gene="PCS0753"
/note="Similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"

CDS

/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BA834176.1"

/db_xref="GI:13360212"
/translation="MNLEHYOAKOLFARVGLPAPVGYCTTPREAEASAKIGAGPMV
VKCVHAGGRKAGVGVVNSKEDIRAFANWLGKRLVYOTDANGOPVMOIIEVAT
DIAEYLGAVNDRSSRRVFMASTEGVEIEVAETPILIKRVLDPLGPMYPO
RELAFLGLEKLVDOOFTKIFMGLATFLERDLALEINLVLTITKGGDLIDCKRGA
DGNALFPQDIREMROSDOPRAOAAQEWLVVLDNIGCVNAGAGLGMTMTIV
KLHGCEPANFLDGGCAFTKRYVEARFIIISDKVKAIVNVTGIVKCDLADGIIIG
AVAEVGVNVVYVRLGNNALCKKLADSGINIIAKGLTDAAGVVAAVECK"

```

VERSION      AE005258.1  GI:12513751
KEYWORDS     Escherichia coli O157:H7 EDL933.
SOURCE       Escherichia coli O157:H7 EDL933
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
              Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
              Postfal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
              Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
              Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
              Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
              Welch, R.A. and Blattner, F.R.
              Genome sequence of enterohemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
TITLE        Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
              Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
              Postfal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
              Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
              Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
              Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
              Welch, R.A. and Blattner, F.R.
              Direct Submission
              Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES     Location/Qualifiers
              1..13501
               /organism="Escherichia coli O157:H7 EDL933"
               /strain="EDL933"
               /serotype="O157:H7"
               /db_xref="taxon:155864"
               /note="enterohemorrhagic"
misc-feature  <..7576
               /note="O-island #36; Region of the EDL933 chromosome not
               homologous to E. coli K-12 MG1655; Cyclic prophage
               CP-933K; includes one copy of the 13 bp direct repeat that
               flanks the prophage"
               66..665
               /gene="20982"
               /note="20981"
               66..665
               /gene="lomK"
               /function="putative membrane; Other or unknown (Phage or
               Prophage Related)"
               /note="Residues 1 to 199 of 199 are 72.86 pct identical to
               residues 1 to 199 of 199 from Genpept 118 :
               g117532789|gb|AF63231.1|AF151091_2 (AF151091) lom
               [prophage P-Elba]"
               /codon_start=1
               /transl_table=11
               /product="putative outer membrane protein of prophage
               CP-933K"
               /protein_id="AAG55137.1"
               /db_xref="GI:12513752"
               /translation="MKKCAATLISAICIAVSPVASEHOSTLSAGYLAHSTAPG
               SDLINCINVKRYEFTDITGLITSEFYANAEDQKTHSDTRHEDYVANKRPSWAG
               PSYVNEWFSAVMAGVAYSRVTSFGDYERVDNRKRHDLVLTGSDARISNTSLAW
               GAGVONPRTESVAVDAVEKSGDMRTDGLVGVYKRF"
               719..2041
               /gene="20982"
               /note="20982"
               /function="putative structure; Structural component (Phage
               or Prophage Related)"
               /note="Residues 164 to 440 of 440 are 68.79 pct identical
               to residues 381 to 645 of 645 from Genpept 118 :
               g114585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
               tail fiber protein [Bacteriophage 933W]"
               /codon_start=1
               /transl_table=11
               /product="putative tail component of prophage CP-933K"

```

```

               /protein_id="AAG55138.1"
               /db_xref="GI:12513753"
               /translation="MGMAVVOISGVTKDQACKPIQNC"ITOLKARNSTXVVVTVASE
               NPEAGRTSMDEYGCOTSYTLTVBGFPPSHAGTISVYDSQPGTILNDLIGAMTBDAR
               PALRRFEQWVEARHAHEAKKNAGEESTASRNAGISASAKASANAADPSADADASE
               SAROAAESAASAKSEASSSASAAKASLSADVADELISKTASTAGNARDA
               TSTEKARESAASAEOSRTAADAVNRIPVVGPPGPKGEPGAPGPKDQKGE
               RQDTGPAGATGEGRGDTGPAGDQGRDREGETGITGNAQPGKGTGAAGA
               GPQGPKEGAGAGPYGATGPGPKGPGPTQIRFLGPMRIETNSYGMFPCTGALIT
               GELTTPDPAQVQGMFQHLQVRRGDGQWQDQKLDDEVGSDTGRTE"
               2004..2232
               /gene="20984"
               /note="20984"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"
               /note="Residues 1 to 102 of 102 are 98.03 pct identical to
               residues 28 to 129 of 129 from Genpept 118 :
               g114585437|gb|AAD25465.1|AF125520_60 (AF125520)
               hypothetical protein [Bacteriophage 933W]"
               /codon_start=1
               /transl_table=11
               /product="unknown protein encoded by prophage CP-933K"
               /protein_id="AAG55139.1"
               /db_xref="GI:12513754"
               /translation="MKWAVIOAFLENDNMIILKMQSLGCGCKHDCNGRSTQALR
               LGPADILESDENGITPEODRVITQVVILDAKKKQIQCVVRLQILRADGTWENTGCMK
               "
               2489..3469
               /gene="20985"
               /note="20985"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"
               /note="Residues 13 to 326 of 326 are 55.73 pct identical
               to residues 18 to 331 of 336 from Genpept 118 :
               g116960367|gb|AF33527.1 (AF170176) hypothetical protein
               predicted by Glimmer [Salmonella typhimurium LT2]"
               /codon_start=1
               /transl_table=11
               /product="unknown protein encoded by prophage CP-933K"
               /protein_id="AAG55140.1"
               /db_xref="GI:12513755"
               /translation="MSPIRTFTHNSVNIQSSPCQTVSPFAGEYELKVIDEXTPIIF
               KLPEDPERRKIKDEVPVNTKQHPYLDNVTNARIIESDRMIGLIFVGDSVNVKTAES
               KLPEPENNMYIYREVDVDFSMYDKLSDIYHDITCSORLRTTEKRDYELNLILELR
               EISKQSLISMVAKKRNHAMPDFRRLALIKAGEIPRCYVNRKNGISGECYILD
               MDNLTLEKRLGTVAPDGISMBVDRNDNVNIENSATIVYNSNHPALLEGISPMHSKVD
               AHPYDGLGKGVKKYFNFTPLHNYNHCFDTEFNHPNITMNTSOTCSSM"
               3530..4522
               /gene="20986"
               /note="20986"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"
               /note="No significant matches"
               /codon_start=1
               /transl_table=11
               /product="unknown protein encoded by prophage CP-933K"
               /protein_id="AAG55141.1"
               /db_xref="GI:12513756"
               /translation="MKTPSLOSNNFNSAPACYSAPLAPNRAENAYADVIDICKRIPL
               SAADLSNYESVIRAVHDRSRLIDQHVNDIGMTVDALSRSOTFDDASVGIHNK
               VHIGIKRYNEELNEESSVKIDDIOSLTCNEKLEYVGQDPIPIICEAGNDNEEY
               VSFSAPTDQSEMSQWEGLIHETLHHVTSSSPGSDNIELGPTFLARVQDEIG
               MSVPFKYGAPEPERAHLRIPLNALRQDAARHRENERAFERLGTISDRYEASPDFT
               EYSAVSNTGIGTIOQHDPLGLAINDNDANOIQLYHCAPIITFFGVQDKNNQ"
               5350..6231
               /gene="20989"
               /note="20989"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"

```


Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaactgctgttgaagcagcag 25
|||||
Db 179 GGATCTCTCTTTCAGCCAGCAG 203

RESULT 12
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
DEFINITION genome
ACCESSION AE000180.1 GI:1786988
VERSION AE000180.1
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
9278503
9278503

REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

REFERENCE 4
AUTHORS This sequence was determined by the E. coli Genome Project at the
TITLE University of Wisconsin-Madison (Frederick R. Blattner, director).
JOURNAL Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of

FEATURES labels.
names.
source This should allow them to be searched for in Entrez as gene
names.

repeat_region
Location/Qualifiers
1..11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1..130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"

gene
complement(147..1430)
/gene="ybhC"
/note="b0772"
complement(147..1430)
/gene="ybhC"
/function="orf; Not classified"

CDS
/note="f427: 98 pct identical to fragment YBHC_ECOLI
SW:P46130 (300 aa) but contains 127 additional C-terminal
residues"

/codon_start=1
/transl_table=11
/product="putative peclintesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786988"

/translation="MMPFYSRLAALAFGYTLTACSSPPPOQPSDQAPETSSRPI
LSAKENQPDQHYFASLTPGAKANPSTTLPAQDFVYVAGTGGVHTHTIOAYD
MALIKRTNRKQIYAVMPGEYQGVYVPAAPGGITLTGTEKPIDVIGISLGGSPA
DMRHVDNPKRQMPKPYAMVYDSCQKRSIDSGYLSAVFWSQNNGLQLOLTLTENT
LSDVDAGNHPVALRTDQDOYINNVNLLGRQNTFFVNSGVONRLERNRPTLT
NSYIEDGVADIVSARGAVDNTFEFVNSRTOGEAVFAPATLSNITYGFLAVNSRFN
AEGDVEAGLRSIDVDANTNGOVYLRDSAINSGFPAKRWADAVISNRFAGNTGSVD
DNEIQRNLNDTNYNMMEYNRNGVSKVAAAKK"

promoter
complement(1507..1535)
/note="factor Sigma70: predicted +1 start at 806574"

gene
complement(1582..2058)
/gene="ybhB"
/note="b0773"

CDS
/function="orf; Unknown"
/note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"

/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"

/db_xref="GI:1786990"
/translation="MKLISNDLRDGLKLRHRYHNGMGYDGDNISPHLAMPDVPAGTK
SFVVCYDADAPDTGGMWVWVNPADTRVLPQFGSLVAMPDGVLOTGRTDGKGTG
YDGAAPPKQETHRVYFTVHALDIERIDVDGASGAMGVFNHPSLASITAMFS"
complement(2108..2124)
/note="central position to predicted promoter: 85.5"
bound_moiety="Rhas predicted site"

gene
complement(2117..3406)
/gene="b10A"

/note="b0774"
complement(2117..3406)
/gene="b10A"

/EC_number="2.6.1.62"

/function="enzyme; Biosynthesis of cofactors, carriers:

Biotin"

/note="f429: 100 pct identical to BIOA_ECOLI SW: P12995"

/codon_start=1

/transl_table=11

/product="7,8-diaminopelargonic acid synthetase"

/protein_id="AAC73861.1"

/db_xref="GI:1786991"

/translation="MTTDDLAFFDQRIHIMPTYSPLPYVVSABGCELLISDGR
LVDKMSWMAIHGYNHPOILAAKSOIDAMSHVFGITIHAPAEICRKLVTMPDP
LCEVFLADSGVAIVEAMKALQYWAQGEAROFLEFRNGHDTGASGVDDPNS
MHSLSMKYILPENLFAAPQSMGEMERDMVGAARLMAHREIAAVITTEPVOGGS
GMRYVHPFWLKRIRKICDRGILLILADEIATGRTGRLFRCEHAETAPDILIGKAL
TGGITMTLSATLTITREVAETISNGEAGCFMGPFTMGPNLACAAANSLATLESQDQO

```

SPVVTGVDPAATGSGMHWVYVNLPAATVTLPOGFGSSGLVAMPDGLQTRTDFGKNG
YDGAAPKPGCTHRTITVTHALDLERIDVDGASGAMYGFWVHFHSLASATAMPAS"
stem_loop
5583. .5605
5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgacagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGACGACG 203

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS A93679 Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source
1.5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141.1156
/standard_name="BIOF RBS"
1154.2308
/gene="BIOF"
1154.2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMOEKINALDARRADALRRRYVAOCAGRMVADDDROYLNF
SSNDYGLSHHPQITRWOOGAEOFGTSGSGSHVSGYSYVHQALEPEELAEMLYSRA
LLEISGFANQAVIAAMAKEDRIADRUASHLEDAASLFSQLEFRFANHYTHLAR
LLASPCGQGVVTEGVFSMDGSAPLAELQOYTOQHNCMLWDADAHGTVGEQKRG
SCMLQKVPPELLVTFGKFGVSGAVALCSYADVLLQFARLLYSTSMPPAQOAI.
RASLAVIRSDGABREKRLAALITRFAGVODI.FFLADSCSAICPLIAGDNRALQI.
AEKTRQGCWVTARIPPTVYAGTARLRLLTAAHEMQDIDRLLEVLHGNG"
RBS 3030.3045
/standard_name="BIOF RBS"
3043.3753
/gene="BIOF"
3043.3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental

```

```

/product="DNR SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEVGTDTFVKGTVASCALLQAAKACRYRAGKPPVASC
EKPEBRLNSDALALQNSLOLDYATVNPYFAPPTSPHIIISAQGPRTSLVMSAG
LRALQOAWVILVEGAGGWFPLESDTFEFADWVTEQQLVLLVGVKIGCJNHAKLTA
QYIQAAGLTLAGVANDVTPPGKRRAEYVTTLRMIPAPLLGEIFWLAENPENMATGK
YINLAFAVDASTIGFTSRL"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgacagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGACGACG 203

RESULT 10
ARI01809 5872 bp DNA PAT 14-FEB-2001
LOCUS ARI01809 Sequence 1 from patent US 6083712.
DEFINITION ARI01809
ACCESSION ARI01809
VERSION ARI01809.1 GI:12812607
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
LONZA AG (CH)
FEATURES
source
1.5872
/organism="unknown"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgacagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGACGACG 203

RESULT 11
ARI01810 5872 bp DNA PAT 14-FEB-2001
LOCUS ARI01810 Sequence 6 from patent US 6083712.
DEFINITION ARI01810
ACCESSION ARI01810
VERSION ARI01810.1 GI:12812608
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
LONZA AG (CH)
FEATURES
source
1.5872
/organism="unknown"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;

```


SOURCE
ORGANISM Escherichia coli (strain K-12) DNA.
Escherichia coli
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.
REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and Johnson,O.
TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted
JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
FEATURES
source
location/Qualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AA23513.1"
/db_xref="GI:455168"
/translation="MKLISNDLRDGLPHRHVNGMGDGDINISPLAMDVDYPAGTKSFVATCYDDPAPGSGMWHVYVNLPAFTFVLPDGFSGSLVAMPDGVLPDRTEGKTGYDGAAPKGETHRVIFTVHALDIERIDVDEGASAAVCGFVNHPSILASASITAMFS"
/gene="biob"
complement(633..1925)
CDS
/gene="biob"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA23514.1"
/db_xref="GI:457106"
/translation="MTTDDLAFDORHIIHPYTSPLPYVVAEGCELLSDRRLVGDSWMAIHGYNHNPOLNAARKSVIDAMSHVMEGITHHAIELCRILVMSGINALECVLADSGSAVEVAEMKMALOYWAKGEARQRLTFNNGYHGDTEGAMSVCDPDN SMHSLMGVLPENLFEAPAPQSRMDGEMDERDMGFARLMAHREHETAAVITIEIYOGA GGMRYHPHWLKRIRKICDREGILLADRELATGFGTRGKLFACENAEIADIDILCKA LTGQMTLSATLTREVAERTISNGEACGPEHGTGPMGNPLACAAANASLAIIESGIMO OOVADIIVOLREOLAPARDAENVADYVCAIGVETTHPVNNAALOKFVEGVNIR PCKLIYMPYITILPQOLQRLTRAANRAVODETFEQ"
2012..3052
/gene="biob"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA23515.1"
/db_xref="GI:145425"
/translation="MAHRPRMILSQVTELEFKPLDLLPEVQOVIRHGFDPPOVYST LLSIKTGACPEDCKYCPQTSRYKTGLAEALMEVEVLISARAKAGSGTRPGMAGV KNPHEDMPLQEMOVGVAMGLEACMTGCTISESQAORLANAGDIYVHNLDTSPEF VGNITTRTYOERLDLTLEKVRDAGIVCSGIVGLDETVKDRAGLLDLQLANLPTPEF VPAINLVKKGCTPLANDVDADFTRTIAVARIMPSTYVRLSAGREOMNQTQAMC PMAGANSIFYGCKLITTPNEEDKDLQLEFRKQLNPQTVAVLAGDNEQOORLEOALMT PTDDEYYNAAL"
3049..4203
/gene="biob"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA23516.1"
/db_xref="GI:145426"
/translation="MSWOKINAAIDARAAALRRRYVACAGRWLVADRYOLNF SSNDYGLSHHPOIIRAMQOQAEORGTSGSGSHVSGISVHQALEELIEALWLVYSRA LFLISGFANQAVIAMAKEDRIADRLSLASLSPLSOLRRRAHNDVPHILAR LIASPCGOOMVYTEGVSMDSAPLAEIQLQVTOQHNGWLMVDAHGTVIGEGRG

gene
CDS
/gene="biob"
/codon_start=1
/transl_table=1
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVKNQALAAFGRAAHYEQHDLQROSDALIAMLPQRKYT HVIDACGCPKMSRMRERHRAQVYALDLSPPMIVQARQDAADHYLAGIESTPLATA TFDLASNLAVQWGMNSTALRELYRVVRPKGVAFPTLVGSGSLPERHQAQVADVER HANRFLPDEIEQSLGVVYOHNIQPTITLMPDALSMSRLGATGATLHREGDRPRL TRSOLRIOLAMPQOOGRRYPLVHPLVGIARE"
4938..5597
/gene="biob"
4938..5597
/gene="biob"
/codon_start=1
/transl_table=1
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVGTDTENGKTVASCALLQAAKRAQYRTAGKYPVAGS EKRPBELRNSDALAQRNSLOLDYATVAPYFAPTSPIISAGEGRPISLVASG LRALFEOADWVLEAGAGWFTPLSDTFPFADWTOEOLPVLLVGVKIGCINHAALTA OVTOHAGLTLAGVANDVTPPKRIHAEVYTLTRMIPRCWRSRGLOKIKMQPES T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.
Query Match 100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ggaatcgctgttggaagcgacgacg 25
Db 2074 GGATCTGCTGTTTGAAGCGACGACG 2098
RESULT 6
A38246
LOCUS A38246 5872 bp DNA
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
COMMENT Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
FEATURES
source
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"


```

71 17.6 70.4 1745 1 MCPBE2
72 17.6 70.4 2540 8 ORCHOREFS
73 17.6 70.4 2813 8 CHOOOREFS
74 17.6 70.4 3429 10 AB029919
75 17.6 70.4 3561 6 AR136150
76 17.6 70.4 3561 6 AR143947
77 17.6 70.4 3561 6 192577
78 17.6 70.4 3710 10 MM0010949
79 17.6 70.4 3998 1 HST06325
80 17.6 70.4 4176 9 HST06325
81 17.6 70.4 5141 9 AF262018
82 17.6 70.4 5552 14 AF128244
83 17.6 70.4 5552 14 AF128244
84 17.6 70.4 5552 14 AF128247
85 17.6 70.4 5552 14 AF128248
86 17.6 70.4 5552 14 AF128249
87 17.6 70.4 5552 14 AF128250
88 17.6 70.4 5552 14 AF128251
89 17.6 70.4 5552 14 AF128252
90 17.6 70.4 5552 14 AF128253
91 17.6 70.4 5597 14 MEAL3
92 17.6 70.4 5597 14 PEVL1
93 17.6 70.4 6597 14 PEVL2
94 17.6 70.4 6639 14 MEALPRO
95 17.6 70.4 6643 14 AB052820
96 17.6 70.4 6830 1 AF198617
97 17.6 70.4 7850 3 BMOTRAS1
98 17.6 70.4 10491 14 MEANC
99 17.6 70.4 10492 14 MEANC
100 17.6 70.4 15727 1 AE001224

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 1041)
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Source
1. 1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ggaatcgtctgttgaagcgacgacg 25
Db 63 GGATCTGCTGTTGAAGCGACGACG 87

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

X66283 M. clara pla
X64615 Oenothera b
X64616 Oenothera o
AB029919 Mus muscu
AR136150 Sequence
AR143947 Sequence
192577 Sequence 12
AJ010949 Mus muscu
AF006325 Lactococc
X54637 Human LYK2
AF262018 Homo sapi
AF128244 Measles v
AF128245 Measles v
AF128247 Measles v
AF128248 Measles v
AF128249 Measles v
AF128250 Measles v
AF128251 Measles v
AF128252 Measles v
AF128253 Measles v
D37775 Measles vit
D37776 Subacute sc
D37776 Subacute sc
M20865 Measles vit
AB052820 Measles v
AF198617 Aeromonas
D38414 Siliworm DN
D63926 Measles vit
D63924 Measles vit
AE001224 Trepnema

```

```

ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1041)
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Source
1. 1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ggaatcgtctgttgaagcgacgacg 25
Db 63 GGATCTGCTGTTGAAGCGACGACG 87

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE
AUTHORS Escherichia.
JOURNAL Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
FEATURES
Source
1 (bases 1 to 1084)
Patent: GB 2216530-A 16 11-OCT-1989;
Location/Qualifiers
1. 1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24. 1064
/gene="B10B"
24. 1064
/gene="B10B"
/codon_start=1
/transl_table=11
/protein_id="CA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRRPRTLSQVTELEFKPLLDLLEAQQVHROHDPVOYST
LISRTGACPEDECKTCPOSSRYKTLGAERLMEVEQVLESARKAAGSTRFCGAM
KNPHERMDPYLEOVQVKGAMGLBACMTLGLTSSQAOQLANAGLDYNNILPTSPF
VGNITTRTYOERLDITLEKVDAGIKVCSGIVIGETVDRAGLILQILNLTPTSP
VPINMLVVKCTPLANDNDVDAFDFIRIIVARIIMPTSVYRSAGREONNEDTOAMC
FMAGANSIFVYGCXKLITTPNPEDEKDLQLEFRILGNPQNTVILAGDNEQORLDQALMT
PQDEYVNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ggaatcgtctgttgaagcgacgacg 25
Db 86 GGATCTGCTGTTGAAGCGACGACG 110

RESULT 4
LOCUS AF250776 5526 bp DNA
DEFINITION Uncultured bacterium pCoshe2 hypothetical 17.1 kDa protein in

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:56:03 : Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-4
Sequence: 1 gacatcgtgttgaagcagcagcag 25

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_ov:*
21: em_pat:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	6 AR029499	AR029499 Sequence
2	25	100.0	1041	6 AR034916	AR034916 Sequence
3	25	100.0	1084	6 A11530	A11530 B10b gene O
4	25	100.0	5526	1 AF250776	AF250776 Uncultured
5	25	100.0	5793	1 E0810	E0810 Sequence 1
6	25	100.0	5872	6 A38246	A38246 Sequence 1
7	25	100.0	5872	6 A38251	A38251 Sequence 6
8	25	100.0	5872	6 A93674	A93674 Sequence 1
9	25	100.0	5872	6 A93679	A93679 Sequence 6
10	25	100.0	5872	6 AR101809	AR101809 Sequence
11	25	100.0	5872	6 AR101810	AR101810 Sequence
12	25	100.0	11022	1 AE000180	AE000180 Escherich
13	25	100.0	12891	1 AE004192	AE004192 Vibrio ch
14	25	100.0	13501	1 AE005258	AE005258 Escherich
15	25	100.0	297816	1 AP002553	AP002553 Escherich
16	23.4	93.6	1121	6 E00893	E00893 Genomic DNA
17	22.4	89.6	8227	1 AF248314	AF248314 Unculture
18	21	84.0	128	1 E0810B	E0810B Sequence
19	19.8	79.2	80993	3 AC004313	AC004313 Drosophila
20	19.8	79.2	83551	2 AC018150	AC018150 Drosophila
21	19.4	77.6	1041	1 E0810TOR2	E0810TOR2 Sequence
22	19.2	76.8	824	6 AX109589	AX109589 Sequence
23	19.2	76.8	86446	2 AC014412	AC014412 Drosophila
24	19.2	76.8	182525	3 AC008213	AC008213 Drosophila
25	19.2	76.8	197348	3 AC008028	AC008028 Drosophila
26	19.2	76.8	239171	3 AE003765	AE003765 Drosophila
27	18.8	75.2	2472	10 AF052042	AF052042 Rattus no
28	18.8	75.2	8371	1 AF087452	AF087452 Bacillus
29	18.8	75.2	16147	3 DME243916	DME243916 Drosophila
30	18.8	75.2	190673	3 AC008357	AC008357 Drosophila
31	18.8	75.2	191736	3 AC025411	AC025411 Homo sapi
32	18.8	75.2	193594	2 AL592563	AL592563 Homo sapi
33	18.8	75.2	221888	3 AE003686	AE003686 Drosophila
34	18.6	74.4	1610	14 AB014726	AB014726 Chuzan vi
35	18.6	74.4	2191	9 AF188738	AF188738 Homo sapi
36	18.6	74.4	4501	1 AFAPBHT1	AFAPBHT1 Alcaigenes
37	18.6	74.4	6572	10 AB030242	AB030242 Mus muscu
38	18.6	74.4	10529	1 AE005138	AE005138 Halobacte
39	18.6	74.4	11208	1 HHRNAPO1	HHRNAPO1 Human DNA
40	18.6	74.4	25042	9 HS380A1	HS380A1 Human DNA
41	18.6	74.4	70326	2 AC068957	AC068957 Homo sapi
42	18.6	74.4	150516	2 AC022946	AC022946 Homo sapi
43	18.6	74.4	151696	2 AF240629	AF240629 Homo sapi
44	18.6	74.4	154279	2 AL138781	AL138781 Homo sapi
45	18.6	74.4	170896	2 AC011010	AC011010 Homo sapi
46	18.6	74.4	190937	2 AP000957	AP000957 Homo sapi
47	18.6	74.4	202177	2 AC016821	AC016821 Homo sapi
48	18.6	74.4	256073	2 AE006464	AE006464 Homo sapi
49	18.6	74.4	340000	2 AP001679	AP001679 Homo sapi
50	18.2	72.8	771	2 AC059196	AC059196 Giardia i
51	18.2	72.8	838	14 AF038279	AF038279 Influenza
52	18.2	72.8	195418	2 AC023524	AC023524 Homo sapi
53	18.2	72.8	198084	2 AC019339	AC019339 Homo sapi
54	18.2	72.8	215126	2 AC083916	AC083916 Mus muscu
55	17.8	71.2	2091	14 AF325487	AF325487 Rabies vi
56	17.8	71.2	5986	2 AC013901	AC013901 Drosophila
57	17.8	71.2	7215	1 SNABR1	SNABR1 Drosophila
58	17.8	71.2	126347	9 HS58A9	HS58A9 Human DNA
59	17.8	71.2	129577	9 AC004659	AC004659 Homo sapi
60	17.8	71.2	168068	9 AL445668	AL445668 Human DNA
61	17.8	71.2	174752	2 AC091624	AC091624 Canis fam
62	17.8	71.2	198456	2 AC003115	AC003115 Homo sapi
63	17.8	71.2	302350	3 AE003485	AE003485 Drosophila
64	17.6	70.4	716	2 AC051152	AC051152 Giardia i
65	17.6	70.4	817	2 AC088150	AC088150 Giardia i
66	17.6	70.4	919	2 AC051207	AC051207 Giardia i
67	17.6	70.4	943	2 AC056463	AC056463 Giardia i
68	17.6	70.4	952	4 AF100454	AF100454 Sus scrofa
69	17.6	70.4	1005	2 AC056672	AC056672 Giardia i
70	17.6	70.4	1684	3 TBMTAT16	X56764 T. brucei MR

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgtctgactctgtgt 18
 |||
 Db 249 GTTGTGATCTGCTGTT 266

RESULT 45

FR0044714 177 bp DNA GSS 22-OCT-1999
 LOCUS Fugu rubripes GSS sequence, clone 192G14e6, genomic survey
 DEFINITION sequence.

ACCESSION AL132206.1 GI:6114152
 VERSION GSS; genome survey sequence.
 KEYWORDS Takifugu rubripes.
 SOURCE Takifugu rubripes.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 177)
 Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
 Umraniya,Y., Williams,G. and Brenner,S.

TITLE Direct Submission
 JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
 Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

COMMENT biohelpehamp.mrc.ac.uk
 Vector: pBluescript II KS
 V-type: phagemid

COMMENT

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES

source location/Qualifiers
 1..177
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 192G14"
 /clone="192G14e6"
 BASE COUNT 33 a 41 c 50 g 44 t 9 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 177;
 Best Local Similarity 90.5%; Pred. No. 2e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgcgtgactctgtgtgaag 23
 |
 Db 110 TCTGTGAATCTGCTGTTGAAG 130

Search completed: December 26, 2001, 12:11:22
 Job time: 6444 sec

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BU0-apa-c-11-0-UI"
/clone_lib="UI-R-BU0"

```

Query Match	72.0%;	Score 18;	DB 11;	Length 487
Best Local Similarity	100.0%;	Pred. No. 1.7e+03;		

PT773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Ronaldo.

BASE COUNT 158 a 194 c 230 g 103 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 685;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tctgtgattctgtctgttgaagc 25
||||| ||||| ||||| ||||| |||||

Db 459 TGCAGGCTCTGCTTTGATGCC 437

RESULT 41
CNS0564D 938 bp DNA GSS 26-JUL-2000
LOCUS CNS0564D/c Tetraodon nigroviridis genome survey sequence T3 end of clone
DEFINITION 028E09 of library A from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL322870.1 GI:9555754
VERSION AL322870.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.

1 (bases 1 to 938)
Rost Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W., and Weissbach, J.
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL MEDLINE 20296633
REFERENCE 2 (bases 1 to 938)
Crollius, H.R., Jallion, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., and Weissbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL MEDLINE 20359837
REFERENCE 3 (bases 1 to 938)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES
source
1. 938
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="028E09"
/clone_lib="A"
/note="Genoscope sequence ID : C0NA028AC05A1-end : T3"

BASE COUNT 237 a 232 c 192 g 248 t 29 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 938;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gtctgtgattctgtctgttgaagc 24
||||| ||||| ||||| ||||| |||||
Db 353 GTTGTGCTCTGCTTTGAGAC 330

RESULT 42
AK006021/c
LOCUS AK006021/c
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700016D22, full insert sequence.

ACCESSION AK006021
VERSION AK006021.1 GI:12838904
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
clone:1700016D22.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1839)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)
99279253
PUBMED 10349636

2 (bases 1 to 1839)
Carninci, P., Shibata, Y., Hayatsu, N., Suganara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)

JOURNAL MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 1839)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Matsumi, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
11076861
PUBMED 20530913
REFERENCE 4 (bases 1 to 1839)
The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 1839)
Adechi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T., Harada, A., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, A., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

TITLE 'G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from wild tomato (*Lycopersicon pennellii*)
trichomes

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
3 prime sequence.

FEATURES
source Location/Qualifiers
1..628

/organism="Lycopersicon pennellii"
/db_xref="taxon:28526"
/clone="CLP1K17"
/clone_lib="L. pennellii trichome, Cornell University"
/tissue_type="trichome"
/dev_stage="mixed stages"
/lab_host="SOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells highly enriched for trichomes, with minor
contamination by other types of leaf cells."

BASE COUNT 184 a 140 c 122 g 182 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 628;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgcgcgacgcgcgtttgaagc 24
||||| ||||| ||||| ||||| |||||
Db 71 TTGCTGATGATGATGATTTTAAAC 49

RESULT 39
TAG9C05P 650 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 69c05, forward sequence,
genomic survey sequence.
ACCESSION AL457556
VERSION AL457556.1 GI:11858782
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 650)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
Project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
nh@sanger.ac.uk
constructed at the institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

DETAILS of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.

FEATURES
source
1..650
/organism="Trypanosoma brucei"

/strain="TREU927"
/db_xref="taxon:5691"
/clone="69c05"

BASE COUNT 208 a 186 c 173 g 83 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 650;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgcgcgacgcgcgtttgaagc 24
||||| ||||| ||||| ||||| |||||
Db 364 TTGCTTATTCGTCTTTGAAGC 342

RESULT 40
A1325366 685 bp mRNA 23-DEC-1998
LOCUS m129T01.Y1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:464953 5' similar to TR:060947 O60947 MAX-INTERACTING
TRANSCRIPTIOMAL REPRESSOR. ; mRNA sequence.
ACCESSION A1325366
VERSION A1325366.1 GI:4059795
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 685)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacey, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennan, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:278769
This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Possible reversed clone; similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 384.

FEATURES
source

1..685
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:464953"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73P-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 1]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified

FEATURES
source
Location/Qualifiers
1..559
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="162c10"

BASE COUNT
198 a 154 c 92 g 115 t

ORIGIN

Query Match
Best local similarity 72.8%; Score 18.2; DB 13; Length 559;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tctgtgactgtctgtttgaagc 25
1 ||||| ||||| ||||| ||||| |||||

Db 196 ttctgtactgtctgtttgaagc 174

RESULT 36
LOCUS AO936593 596 bp DNA GSS 23-AUG-2000
DEFINITION HSJ41-840 Human Noli clones Homo sapiens genomic, DNA sequence.
ACCESSION AO936593
VERSION AO936593.1 GI:7212971
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 596)
Zabarovskiy, E.R., Gizatulilin, R., Podowski, R.M., Zabarovska, V.V., Xie
L., Muravenko, O.V., Kozirev, S., Petrenko, L., Skobeleva, N., Li, D.,
Protopopov, A., Kashuba, V., Ernberg, I., Winberg, G. and Wahlstedt, C.
Noli clones in the analysis of the human genome
Nucleic Acids Res. 28 (7), 1635-1639 (2000)
20175728
TITLE Contact: Podowski RM
JOURNAL Center for Genomics Research
MEDLINE Karolinska Institute
17177 Stockholm, Sweden
Tel: +46-8-728-6372
Fax: +46-8-337983
Email: Raf.Podowski@cgrr.ki.se
Class: Noli site.
FEATURES
source
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="Human Noli clones"

BASE COUNT
138 a 200 c 182 g 75 t 1 others

ORIGIN

Query Match
Best local similarity 72.8%; Score 18.2; DB 13; Length 596;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtgtgactgtctgtttgaag 23
1 ||||| ||||| ||||| ||||| |||||

Db 488 gtgtgactgtctgtttgaag 466

RESULT 37
LOCUS AO941180 621 bp DNA GSS 27-JAN-2000
DEFINITION Sheared DNA-19C12.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AO941180
VERSION AO941180.1 GI:6764445
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei

REFERENCE
AUTHORS 1 (bases 1 to 621)
El-Sayed, N., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other-GSSs: Sheared DNA-19C12.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ARCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tdb/mdb/tbdb/>.
Seq primer: M13-Forward
Class: shotgun.
Location/Qualifiers
1..621
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone.lib="Sheared DNA-19C12"
/clone.lib="Sheared DNA-19C12"
/note="Vector: pUC18; Site: 1; SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrell, Oxford University
Press, 1999)."

BASE COUNT
235 a 167 c 132 g 87 t

ORIGIN

Query Match
Best local similarity 72.8%; Score 18.2; DB 13; Length 621;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtgtgactgtctgtttgaag 23
1 ||||| ||||| ||||| ||||| |||||

Db 77 gtgtgactgtctgtttgaag 55

RESULT 38
LOCUS AM160263 628 bp mRNA EST 18-MAY-2001
DEFINITION EST290121 L. penneilli trichome, Cornell University Lycopersicon
penneilli cDNA clone cLPTK17 similar to protein kinase, putative,
mRNA sequence.
ACCESSION AM160263
VERSION AM160263.1 GI:6279797
KEYWORDS EST.
SOURCE Lycopersicon penneilli.
ORGANISM Lycopersicon penneilli.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
AUTHORS 1 (bases 1 to 628)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Lakey, J., Holt, I.E.,
Liang, F., Hansen, T.S., Upton, J., Ronning, C.M., Craven, M.B., Fujii,
C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin

DEFINITION T. brucei sheared genomic DNA clone 375e11, forward sequence, genomic survey sequence.

ACCESSION AL495591

VERSION AL495591.1 GI:11873033

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 512)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

JOURNAL Direct Submission

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

DETAILS OF T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/Projects/T-brucei/>.

FEATURES

Source

1..532

Location/Qualifiers

/organism="Trypanosoma brucei"

/strain="TREU927"

/db_xref="taxon:5691"

/clone="375e11"

BASE COUNT 162 a 156 c 92 g 122 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 532;

Best Local Similarity 87.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 tgcgcgactcgtcgttgaagcg 25

||||| ||||||| |||||||

Db 318 TTCTGTAAGTCTGTGTAAGCG 296

RESULT 34

LOCUS AZ003494 541 bp DNA 24-FEB-2000

DEFINITION RPT-23-372622.TJ RPT-23 Mus musculus genomic clone RPT-23-372622

ACCESSION AZ003494

VERSION AZ003494.1 GI:7078850

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 541)

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shalman, S., Akhmet, B., Levins, M., Megann, S., Tsengaye, G., Geer, K., Kroll, M., de Jong, P. and Fraser, C. M.

TITLE Mouse BAC End Sequences from Library RPT-23 unpublished (1999)

JOURNAL Other-GSS: RPT-23-372622.TV

COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPT-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 372 row: G column: 22

Seq primer: SP6

Class: BAC ends.

FEATURES

Source

1..541

Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPT-23-372622"

/clone_11b="RPT-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 129 a 124 c 154 g 134 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 541;

Best Local Similarity 87.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gtcgcgactcgtcgttgaag 23

||||| ||||||| |||||||

Db 184 GTTCTGATGATGATGTTGAAG 206

RESULT 35

LOCUS TA162C10P/c 559 bp DNA 13-DEC-2000

DEFINITION T. brucei sheared genomic DNA clone 162c10, forward sequence, genomic survey sequence.

ACCESSION AL472440

VERSION AL472440.1 GI:11837892

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 559)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Direct Submission

JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

DETAILS OF T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/Projects/T-brucei/>.


```

/clone="IMAGE:936819"
/clone_lib="NCL_CGAP_CO3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/notes="vector: p1773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p1773 vector. Library went through one round of
normalization."

```

BASE COUNT	177 a	89 c	71 g	158 t
ORIGIN				
Query Match				
Best Local Similarity	72.8%;	Score 18.2;	DB 10;	Length 495;
Matches 20; Conservative	87.0%;	Pred. No. 1.5e+03;		
	0;	Mismatches 3;	Indels 0;	Gaps 0;

```

Oy      1  gtgtcgtgatactgcctgttgaag  23
         |||||  ||  |||||  ||||
Db      451  GTTCTGGACCTACTGTTTCGAAG  429

```

BASE COUNT	129 a	131 c	107 g	130 t	2 others
ORIGIN	Library may contain a small percentage of host or bacterial contaminants. NOTE: Many clones lack the 5' EcoRI site but contain the upstream polylinker sites."				
Query Match	72.8%; Score 18.2; DB 10; Length 499;				
Best Local Similarity	87.0%; Pred. No. 1.5e+03;				
Matches	20; Conservative	0;	Mismatches	3;	Indels 0; Gaps 0;
QY	2	ttgcctgcatacctctgttgaagc	24		
Db	178	TTGCTGCTTCTGCTTTTTCACGC	200		
RESULT	30				
LOCUS	AZ244099	507 bp	DNA	GSS	15-JUN-2000
DEFINITION	RPCT-23-36K23.TJ RPCT-23 Mus musculus genomic clone RPCT-23-36K23, DNA sequence.				
ACCESSION	AZ244099				
VERSION	AZ244099.1	GI:8557290			
KEYWORDS	GSS.				

RESULT	29
LOCUS	AW702183
DEFINITION	AM702183 499 bp mRNA EST 22-MAY-2000 TGGSTGzzz73b05.y1 TgRk+-tachyzoite cDNA Toxoplasma gondii CDNA clone
ACCESSION	TGGSTGzz73b05.y1 5' - rRNA sequence.
VERSION	AM702183
KEYWORDS	AM702183..1 GI:7586334
SOURCE	EST.
ORGANISM	Toxoplasma gondii. Toxoplasma gondii. Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeria; Sarcocystidae; Toxoplasma. 1 (bases 1 to 499) Hell,J., Manger,I., Marra,M., Sibley L.D., Ajlaka,J.A., Aslett,M.A, Waterston,K.H. and Boothroyd,J. WashU-Merck-Stanford-NIH Toxoplasma EST project Unpublished (1996) Contact: Marra M WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: toxo@watson.wustl.edu Contact David Sibley (toxos@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability Seq primer: -40BP from Gibco High quality sequence stop: 415. Location/Qualifiers 1..499
REFERENCE	
AUTHORS	
JOURNAL	
COMMENT	
FEATURES	
source	

RESULT	30
LOCUS	AZ244099
DEFINITION	AZ244099 507 bp DNA GSS 15-JUN-2000
ACCESSION	RPCT-23-36K23.TJ RPCT-23 Mus musculus genomic clone RPCT-23-36K23,
VERSION	DNA sequence.
KEYWORDS	AZ244099
SOURCE	AZ244099.1 GI:8557290
ORGANISM	GSS.
REFERENCE	house mouse.
AUTHORS	Mus musculus
COMMENT	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 507)
TITLE	Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akintet
JOURNAL	, B., Levins, M., McGann, S., Tesgaye, G., Geer, K., Krol, M., de Jong, P.
COMMENT	Mouse BAC End Sequences from Library RPCT-23
	Unpublished (1999)
	Other_GSS: RPCT-23-36K23.TV
	Contact: Shaying Zhao
	Department of Eukaryotic Genomics
	The Institute for Genomic Research
	9712 Medical Center Dr., Rockville, MD 20850, USA
	Tel: 301 838 0200
	Fax: 301 838 0208
	Email: szhao@tigr.org
	Clones are derived from the mouse BAC library RPCT-23. For BAC
	library availability, please contact Pieter de Jong
	(pieter@jeon.med.buffalo.edu). Clones may be purchased from
	BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
	or from Resea ch Genetics (info@resgen.com). BAC end page:
	http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
	Plate: 36 row: K column: 23
	Seq primer: SP6
	Class: BAC ends.
FEATURES	Location/Qualifiers
source	1..507

```

FEATURES
    source
        location/Qualifiers
            1..499
                /organism="Toxoplasma gondii"
                /strain="RH (type 1)"
                /db_xref="taxon:5811"
                /clone="IGS1273b05.y1"
                /clone_id="TgRH+Tachyzoite cDNA"
                /dev_stage="Tachyzoite"
                /lab_host="SOLR cells"
                /note="Vector: pBluescript SK-; site_1: EcoRI; site_2:
XhoI; cDNA was synthesized from poly mRNA using an
oligo-dT primer containing an XhoI site. Following second
strand synthesis, EcoRI adapters were ligated to the cDNA
and products were size-selected on sephacryl S500. The
cDNAs were ligated to EcoRI/ XhoI prepared lambda ZapII
(Stratagene).. Clones were converted to phagemids by mass
excision using ExAssist helper phage and E.coli SOLR cells
(Stratagene). Insert sizes range from 0.3-3.0 kb. The

```

```

FEATURES
source
    location/Qualifiers
        1. 507
            /organism="Mus musculus"
            /strain="C57Bl/6J"
            /db_xref="taxon:10090"
            /clone="RPC1-23-36K23"
            /clone_11p="RPC1-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
            EcorRI; Site_2: EcorRI; Female C57Bl/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcorI and EcorI Methylase. Size
            selected DNA was cloned into the pBACe3.6 vector at the
            EcorI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT
    145 a 108 c 98 g 156 t
ORIGIN

```

AUTHORS	Hall,N., Bowman,S., Leonard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.F., Rajandream,M.A. and Barrett,B.G.
TITLE	Direct Submission
JOURNAL	Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk
COMMENT	Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, B. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
FEATURES	Email: nelsayed@tigr.org Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/T_brucei/.
SOURCE	Location/Qualifiers 1..464 /organism="Trypanosoma brucei" /strain="TREU927" /db_xref="taxon:5691" /clone="372a07"
BASE COUNT	167 a 132 c 102 g 63 t
ORIGIN	
Query Match	72.8%; Score 18.2; DB 13; Length 464;
Best Local Similarity	87.0%; Pred. No.1.5e+03;
Matches	20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	1 gtctcgcgactcgtcgattgaag 23
Db	39 GTGCTGCATTTGGTGTTGTAG 17
RESULT 28	
LOCUS	AA527258/c
DEFINITION	mRNA sequence.
ACCESSION	AA527258
VERSION	AA527258
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 495) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE	Unpublished (1997)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov
COMMENT	Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arraying: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www.bio.lnl.gov/bdrp/image/image.html Insert length: 1190 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 287. Location/Qualifiers 1..495 /organism="Homo sapiens" /db_xref="taxon:9606"
FEATURES	
SOURCE	

Query Match	72.8%	Score 18.2;	DB 13;	Length 422;
Best Local Similarity	87.0%	Pred. No. 1.4e+03;		
Matches 20; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

RESULT	25			
LOCUS	A1449930/c			
DEFINITION	A1449930	437 bp	EST	09-MAR-1999
ACCESSION	U0818105.x1	Stratagene mouse heart	(#937310)	Mus musculus cDNA clone
VERSION	A1449930	IMAG:603824.3	'	mRNA sequence.
KEYWORDS	A1449930.1	GI:4293448		
EST.				

REFERENCE
AUTHORS
TITLE

(Pages 1 to 437)

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T.,
underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, J., Person,
B., Swallier, T., Gibbons, M., Page, D., Harvey, N., Schurr, R.,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.

The WashU-MCI Mouse EST Project 1999

COMMENT
Contact: Marra M/washu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 393.

```

/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:603824"
/clone_11b="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (Kanamycin resistant)"

```

	BASE COUNT	121 a	76 c	99 g	140 t	Others
	ORIGIN						1 Others
Query Match			72.8%	Score 18.2;	DB 10;	Length 437;	
Best Local Similarity			87.0%;	Pred. No. 1.4e+03;			
Matches 20:	Conservative	0;	Mismatches	3;	Indels	Gaps	0;
2	tctgctgaatcgtgcgtttgaagc	24					
31	TTGCTGGATCTGCCGTTCAGC	9					
b							

RESULT 26
AQ522448

elements were found in this cDNA sequence: 1-22,
>AT_rich#low_complexity
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..303
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UT-M-B21-btu-n-15-0-UT"
/clone_lib="NIH_BMAP_MH12.S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pUT3D-pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_MH12.S1 library is a subcloned library derived
from NIH_BMAP_MH12. NIH_BMAP_MH12 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_LIB=NIH_BMAP_MH12.S1
TAG_TISSUE=hippocampus
TAG_SEQ=TACTC

BASE COUNT
78 a 77 c 60 g 88 t

Query Match 72.8%; Score 18.2; DB 11; Length 303;
Best Local Similarity 87.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttctgagatcgtctgttgaagc 25
|||||
Db 263 TGCTGATCTGCTGCTCAGCG 285

RESULT 22
BG008569/c 345 bp mRNA EST 24-JAN-2001
LOCUS BG008569
DEFINITION PM4-GN0304-291100-001-c07 GN0304 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG008569
VERSION BG008569.1 GI:12453896
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 345)
Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,
Negal,M.A., da Silva,W.J.R., Zago,M.A., Borlin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baita,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 105, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-GN0304-
291100-001-c07&t3=2000-11-29&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 345.

FEATURES
source
Location/Qualifiers
1..345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0304"
/dev_stage="Adult"
/note="Organ: placenta,normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from OESTRES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT
84 a 98 c 89 g 74 t

Query Match 72.8%; Score 18.2; DB 11; Length 345;
Best Local Similarity 87.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttctgagatcgtctgttgaagc 24
|||||
Db 242 TTGCTGCTCTGCTGTTCAAGC 220

RESULT 23
B1141537/c 385 bp mRNA EST 03-JUL-2001
LOCUS B1141537
DEFINITION IP1_53_E10.B1-A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION B1141537
VERSION B1141537.1 GI:14593980
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE
AUTHORS
1 (bases 1 to 385)
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt
L.H.
An EST database for Sorghum: developing preanthesis pannicles
Unpublished (2001)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for high quality sequence is
20. Three-prime sequences, which are obtained with polyTmix or T7
sequencing primer, are presented as the reverse complement.
Seq primer: JEN KEY
High quality sequence stop: 379
POLYA=No.

FEATURES
source
Location/Qualifiers
1..385
/organism="Sorghum bicolor"
/cultivar="PTx623"
/db_xref="taxon:4558"
/clone_lib="Immature pannicle 1 (IP1)"
/note="Organ: Developing preanthesis pannicles; Vector:
pbluescript II SK(-) from Lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; The library was made from polyA RNA in the
cloning vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT
81 a 111 c 83 g 110 t

TITLE	Weissenbach,J.
JOURNAL	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 849)
TITLE	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brotlier,F., Quetier,F., Saurin,W. and Weissenbach,J.
JOURNAL	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 849)
TITLE	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases This sequence is a single read and was generated as part of a large scale clone end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES	Location/Qualifiers
source	1..849
	/organism="Tetraodon nigroviridis"
	/db_xref="taxon:99883"
	/clone="13IG13"
	/clone_11b="G"
	/note="Genoscope sequence ID : C0AG13AD07SP1-end :
	PUC-Orl"
BASE COUNT	233 a 211 c 193 g 208 t 4 others
ORIGIN	
Query Match	73.6% Score 18.4; DB 13; Length 849;
- Best Local Similarity	95.0%; Pred. No. 1.2e+03;
- Matches	19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy	5 ctgcatctactgttgaagc 24
Dd	167 CTGGAGCTGCTTTTGACGC 148
RESULT 15	CNS03713 859 bp DNA GSS 15-MAY-2000
LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION	001G13 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL230736
VERSION	AL230736..1 GI:7889731
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Aclinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 859)
AUTHORS	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 859)
AUTHORS	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brotlier,F., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 859)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases

COMMENT
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
1..859
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="001G13"
/clone_1lb="G"
/note="Genoscope sequence ID : C0Bg001AD07LP1-end : T7"

BASE COUNT
254 a 199 c 178 g 228 t

ORIGIN

Query Match 73.6% Score 18.4; DB 13; Length 859;
Best Local Similarity 95.0%: Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

0y 5 ctgcagctcgtctgttaaac 24
||||| |||||||

Db 242 CTGCAGCTCTCTGTTTCAC 223

RESULT 16
AI605313 85 bp mRNA EST 21-APR-1999
LOCUS AI605313/C
DEFINITION vn87h03.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:1038965 3', mRNA sequence.
AI605313
AI605313.1 GI:4614480
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 85)
Matta,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Sallinger,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:578989
This clone was previously sequenced on the 5' end only, this new data is from the 3' end.
location/Qualifiers
1..85
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1038965"
/clone_1lb="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts"
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTTCGACGACAG 3' -3' adaptor sequence: 5' CTGCAGCTCTCTGTTTCAC 3' -19 t

BASE COUNT
24 a 21 c 19 g 19 t

TITLE
JOURNAL
COMMENT

Bugri, O., Buehl, C.R., Romning, C., Tanksley, S. and Baker, B.
Generations of ESTs from sprouting potato eyes
Unpublished (2000)
Contact: Cathy Romning
The Institute for Genomic Research
For clone info, please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cduan@resgen.com
Seq primer: MJ3F-R.

FEATURES
SOURCE

Location/Qualifiers
1..662
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSTS611"
/clone_lib="CSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."

BASE COUNT 148 a 177 c 142 g 195 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 662;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gtctcgagatcgtctgttgaagc 25
||||||| ||||| |||||
Db 288 GTTGTGAGAGTGTGTTCAAGCG 264

RESULT 11
LOCUS CENS06NLT 1029 bp DNA GSS 17-JUN-2001
DEFINITION T3 end of clone AU00AA015B11 of library AU00A from strain CBS 3082
ACCESSION AL406927
VERSION AL406927.1 GI:12171587
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri.
ORGANISM Saccharomyces kluyveri.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 1029)
Neuveglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F.,
Galliard, C., and Casaregola, S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
FEMS Lett. 487 (1), 56-60 (2000)

JOURNAL MEDLINE
REFERENCE 20584719
AUTHORS

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

3 (bases 1 to 1029)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
amundata, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
SOURCE

Location/Qualifiers
1..1029
/organism="Saccharomyces kluyveri"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU00AA015B11"
/clone_lib="AN00A"
/note="end : T3"
BASE COUNT 328 a 272 c 186 g 242 t 1 others
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 1029;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gtctcgagatcgtctgttgaagc 25
||||||| ||||| |||||
Db 261 GTTGTGAGAGTGTGTTCAAGCG 237

RESULT 12
LOCUS AK015926/c 1329 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4930528G09, full insert sequence.
ACCESSION AK015926
VERSION AK015926.1 GI:12854457
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
clone:4930528G09.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

1 (bases 1 to 1329)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)
99279253
10349636
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

3 (bases 1 to 1329)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Komno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
20530913
11076861
REFERENCE
4 (bases 1 to 1329)

2-day establishment period, seedlings were subjected to a 5 ppm aluminum stress prior to tissue harvest. Plants were grown in an environmental chamber. The tissue, total RNA, and poly(A) RNA were prepared, and a cDNA library was made (Butler and Gustafson) at University of Missouri, Columbia. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson Lab (all other authors).

BASE COUNT 125 a 96 c 134 g 142 t 1 others
ORIGIN

Query Match 75.2% Score 18.8; DB 10; Length 498;
Best Local Similarity 90.9%; Pred. No. 8.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0y 2 ttgcgtgacgtcgttgaag 23
|||||
Db 37 ttgcgtgacgtcgttgaag 16

RESULT 8
AM501847/c 290 bp mRNA EST 01-MAR-2000
LOCUS UI-HP-BR0P-aj2-g-01-0-UI.r1 NIH_MGC_52 Homo sapiens cDNA clone
DEFINITION IMAGE:3075174 5', mRNA sequence.
ACCESSION AM501847
VERSION AM501847.1 GI:7115764
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 290)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: M13 Forward.

FEATURES

Location/Qualifiers

1..290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3075174"
/clone_id="NIH_MGC_52"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(7.4-9.5kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 60 a 108 c 99 g 23 t
ORIGIN

Query Match 74.4% Score 18.6; DB 10; Length 290;
Best Local Similarity 84.0%; Pred. No. 9.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
0y 1 gtgcgtgacgtcgttgaagc 25
|||||

Db 236 gtgcgtgacgtcgttgaagc 212

RESULT 9
AM503005/c 300 bp mRNA EST 01-MAR-2000
LOCUS UI-HP-BR0P-aj2-g-01-0-UI.r1 NIH_MGC_52 Homo sapiens cDNA clone
DEFINITION IMAGE:3076393 5', mRNA sequence.
ACCESSION AM503005
VERSION AM503005.1 GI:7118004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 300)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: M13 Forward.

FEATURES
Location/Qualifiers

1..300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3076393"
/clone_id="NIH_MGC_52"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(7.4-9.5kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 61 a 111 c 102 g 26 t
ORIGIN

Query Match 74.4% Score 18.6; DB 10; Length 300;
Best Local Similarity 84.0%; Pred. No. 9.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

0y 1 gtgcgtgacgtcgttgaagc 25
|||||
Db 236 gtgcgtgacgtcgttgaagc 212

RESULT 10
BG593982/c 662 bp mRNA EST 12-APR-2001
LOCUS EST92660 csts Solanum tuberosum cDNA clone csts611 5' sequence,
DEFINITION mRNA sequence.
ACCESSION BG593982
VERSION BG593982.1 GI:13612122
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Asteridae; eusterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 662)
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A.,

RESULT	3
AM636257	618 bp mRNA EST 26-Apr-2001
LOCUS	b14af04.w1 Blackshear/Sources normalized Xenopus egg library sequence.
DEFINITION	laevis cDNA clone PBX0044F04 5', mRNA sequence.
ACCESSION	AM636257
VERSION	AM636257.1 GI:7393338
KEYWORDS	EST.
SOURCE	African clawed frog.
ORGANISM	Xenopus laevis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopus; Xenopodinae; Xenopus.
AUTHORS	1 (bases 1 to 618) Blackshear,P.J., Lai,M.S., Thorn,J.M., Kennington,E.A., Staffa,N.G., Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman,J.W., Bonaldo,M.F. and Soares,M.B. The NIHES Xenopus maternal EST project: interim analysis of the first 13,879 ESTs from unfertilized eggs Gene 267 (1), 71-87 (2001)
TITLE	21211403
JOURNAL MEDLINE COMMENT	Contact: Perry J. Blackshear Office of Clinical Research and Laboratory of Signal Transduction National Institute of Environmental Health Sciences A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709, USA Tel: 919 541-4899 Fax: 919 541-4571 Email: blackc009niehs.nih.gov Clone is available through Research Genetics, Inc., 2130 Memorial Parkway, Huntsville, AL 35901 phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email cdna@resgen.com DNA sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC).
PCR primers	FORWARD: TGTAAAGCAGCGCCAG BACKWARD: CAGCAACACTATGACC Plate: 0044 row: F column: 04 Seq primer: T7 primer.
FEATURES	Location/Qualifiers 1..618 /organism="Xenopus laevis" /db_xref="Taxon:8355" /clone="PBX0044F04" /clone_lib="Blackshear/Soares normalized Xenopus egg library" /sex="female" /tissue_type="unfertilized egg" /cell_type="unfertilized egg" /dev_stage="unfertilized egg" /lab_host="DH10B" /note="Vector: pT7/T3-Pac; Site.1: EcoRI; site.2: NotI; polyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT7/T3-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. "Normalization and subtraction: two approaches to facilitate gene discovery", genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adaptors, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7/T3-Pac vector. The library contained approximately 7.2 x 10 ⁵ recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT	176 a 106 c 107 g 229 t
ORIGIN	
Query Match	80.84; Score 20.2; DB 10; Length 618;
Best Local Similarity	88.04; Prd. No. 2,3e+02;
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY	1	gttgctgactcgtctttaaag	25
LOCUS			
Db	325	GTTGCTGAATCGTCAATTGAACG	349
RESULT	4		
CNS04KH2		888 bp	DNA
CNS04KH2			GSS
DEFINITION		Tetraodon nigroviridis genome survey sequence T7 end of clone 116D03 of library G from Tetraodon nigroviridis, genomic survey sequence.	21-MAY-2000
ACCESSION		AL294815.1	GI:8033395
KEYWORDS		GSS: genome survey sequence.	
SOURCE		Tetraodon nigroviridis.	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.	
REFERENCE		1 (bases 1 to 888)	
AUTHORS		Roeest-Crollius,H., Jalllon,O., Dasilva,C., Fizesmes,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
TITLE		Unpublished	
JOURNAL		2 (bases 1 to 888)	
REFERENCE		Roeest-Crollius,H., Jalllon,O., Dasilva,C., Bonneau,L., Fisher,C., Bernot,A., Fizesmes,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	
AUTHORS		Unpublished	
TITLE		3 (bases 1 to 888)	
JOURNAL		Genoscope.	
REFERENCE		Direct Submission	
AUTHORS		Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases	
TITLE		This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .	
COMMENT		Location/Qualifiers	
FEATURES		1..888	
source		/organism="Tetraodon nigroviridis"	
		/db_xref="taxon:99883"	
		/clone="116D03"	
		/clone.lib="HG"	
		/note="Genoscope sequence ID : COBG116CB02LP1-end : T7"	
BASE COUNT		129 a	276 c
ORIGIN		276 g	187 t
			20 others
Query Match		77.6%;	Score 19.4; DB 13; Left^v. 888;
Best Local Similarity		95.2%;	Pred. No. 5e+02;
Matches		20; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	3	tgcctgatctgcctgttgaag	23
LOCUS			
Db	331	TGCTGCATCTGCTGTGGAYG	351
RESULT	5		
CNS002W7		1101 bp	DNA
CNS002W7			GSS
DEFINITION		Drosophila melanogaster genome survey sequence TE73 end of BAC # BACRP07D12 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	03-JUN-1999
ACCESSION		AL063545	GI:4941301
VERSION		AL063545.1	
KEYWORDS		GSS.	
SOURCE		fruit fly.	

ALIGNMENTS

86	17.6	70.4	304	10	AA320304	EST22640
87	17.6	70.4	311	11	BT770854	RC1-IT001
88	17.6	70.4	320	10	AM820133	QVO-ST029
89	17.6	70.4	330	11	T19221	Tes
90	17.6	70.4	330	11	T19221	246117 HSC07A021 n
91	17.6	70.4	340	11	243924	243924 HSC10G091 n
92	17.6	70.4	347	10	AA304988	EST176229
93	17.6	70.4	359	10	AT329772	B8D03ne.r
94	17.6	70.4	375	10	AA345079	EST51063
95	17.6	70.4	377	11	BS982632	IL5-CN006
96	17.6	70.4	393	13	AO192143	HS-2251.B
97	17.6	70.4	395	13	B83479	RPC111-15C2
98	17.6	70.4	401	11	R11836	YF53007.r1
99	17.6	70.4	405	10	AA044684	2K5C11.r
100	17.6	70.4	405	11	BC483062	602502882

RESULT 1
CNS051ED/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
051011 of library A from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION
AL316750.1 GI:9549634
VERSION
GSS: genome survey sequence.
KEYWORDS
Tetraodon nigroviridis.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE
1 (bases 1 to 610)
Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fitzames C., Wincker P., Brottier P., Quetier F.,
Saurin W. and Weissenbach J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
2 (bases 1 to 610)
Crolius H.R., Jallion O., Dasilva C., Ozouf-Costaz C., Fitzames C.,
Bernot A., Bouneau L., Billault A., Quetier F., Saurin W.,
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
3 (bases 1 to 610)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
Location/Qualifiers
1. 610
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="051011"
/clone_1lb="A"

BASE COUNT
ORIGIN
124 a 159 c 162 g 139 t 26 others
/note="Genoscope sequence ID : C0AA051AH06C1-end : T7"

Query Match 83.2%; Score 20.8; DB 13; Length 610;
Best Local Similarity 91.7%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ctgtcgtatctcgtctgttgaagc 25
|||||
Db 338 TGGCTGATCTGCAGTTTGAACG 315

RESULT 2
AM766819/c
LOCUS
DEFINITION
AM766819 465 bp mRNA EST 16-FEB-2001
da71d10.x1 Harland stage 19-23 Xenopus laevis cDNA clone
IMAGE:3200371.3', mRNA sequence.
ACCESSION
AM766819
VERSION
AM766819.1 GI:7698812
KEYWORDS
EST.
SOURCE
African clawed frog.
Xenopus laevis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 465)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person,
B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterson, R. and Wilson, R.,
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by R. Harland, PhD (University of California,
Berkeley)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at:
<http://imgl.llnl.gov/image/html/resources.shtml>
Seq primer: 400P from Gibco
High quality sequence stop: 462.
Location/Qualifiers
1. 465
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3200371"
/clone_1lb="Harland stage 19-23"
/tissue_type="neurula"
/dev_stage="stage 19-23"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCS107 (custom); Site_1: NotI; Site_2: SalI
: cDNA made by oligo-dT priming. Library constructed by
Dr. Francesca Mariani in the laboratory of R. Harland,
Ph.D. (University of California, Berkeley). References:
XBF-2 is a transcriptional repressor that converts
ectoderm into neural tissue. Mariani, FV, Harland, RM,
Development. 1998 Dec;125(24):5019-31. PMID: 9811586; UI:
99030283; Use of large-scale expression cloning screens in
the xenopus laevis tadpole to identify gene function.
Gramer TC, Liu KJ, Mariani FV, Harland RM, Dev Biol.
2000 Dec 15;228(2):197-210. PMID: 11112324; UI: 20564075;
Note: This is a Xenopus Gene Collection (XGC) library."

FEATURES
source

BASE COUNT
ORIGIN
141 a 93 c 76 g 155 t
/note="Genoscope sequence ID : C0AA051AH06C1-end : T7"

Query Match 80.8%; Score 20.2; DB 10; Length 465;
Best Local Similarity 88.0%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 gtgtcgtatctcgtctgttgaagc 25
|||||
Db 28 GTTCTGATCTGCAGTTTGAACG 4

Thu Dec 27 16:12:13 2001

us-09-396-196f-3.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:11 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-3
Perfect score: 25
Sequence: 1 gttcgtgactcgtcgttgaagcgc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_estc2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_lin:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.8	83.2	610	13	CNS051ED AL136750 Tetradon
C 2	20.2	80.8	465	10	AM766819 da71d10.x
C 3	20.2	80.8	618	10	AM636257 bl44f04.w
C 4	19.4	77.6	888	13	CNS04K12 AL1294815 Tetradon
C 5	19.2	76.8	1101	13	CNS002M7 AL063545 Drosophi1
C 6	18.8	75.2	412	11	BC233131 daa32f11.
C 7	18.8	75.2	498	10	BE586432 WHE504.D0
C 8	18.6	74.4	290	10	AM501847 UI-HF-BRO
C 9	18.6	74.4	300	10	AM503005 UI-HF-BRO
C 10	18.6	74.4	662	11	BG593982 EST492660
C 11	18.6	74.4	1029	13	CNS06NLT AL406927 T3 end of
C 12	18.6	74.4	1329	12	AK015926 Mus muscu

C 13	18.4	73.6	594	13	FR0036852
C 14	18.4	73.6	849	13	CNS02ESU
C 15	18.4	73.6	859	13	CNS03713
C 16	18.2	72.8	85	10	AI605313
C 17	18.2	72.8	101	10	AA624863
C 18	18.2	72.8	106	10	AI642276
C 19	18.2	72.8	222	11	BI141573
C 20	18.2	72.8	293	13	AZ214122
C 21	18.2	72.8	303	13	BE990318
C 22	18.2	72.8	345	11	BG008569
C 23	18.2	72.8	385	11	BI141537
C 24	18.2	72.8	422	13	AZ412964
C 25	18.2	72.8	437	13	AO522448
C 26	18.2	72.8	437	13	AO522448
C 27	18.2	72.8	464	13	TA372A07Q
C 28	18.2	72.8	495	10	AA527258
C 29	18.2	72.8	507	13	AW702183
C 30	18.2	72.8	517	10	AZ244099
C 31	18.2	72.8	521	10	BE656252
C 32	18.2	72.8	532	13	BE443324
C 33	18.2	72.8	532	13	TA375E11P
C 34	18.2	72.8	541	13	AZ003494
C 35	18.2	72.8	559	13	TA162C10P
C 36	18.2	72.8	596	13	AO936593
C 37	18.2	72.8	621	13	AO941180
C 38	18.2	72.8	628	10	AM160263
C 39	18.2	72.8	650	13	TA69C05P
C 40	18.2	72.8	685	10	AI325366
C 41	18.2	72.8	938	13	CNS0564D
C 42	18.2	72.8	1839	12	AK006021
C 43	18.2	72.8	403	10	BE095496
C 44	18.2	72.8	487	11	BE428015
C 45	17.8	71.2	177	13	FR0044714
C 46	17.8	71.2	240	10	AU074974
C 47	17.8	71.2	295	11	BE440174
C 48	17.8	71.2	296	10	HB190815
C 49	17.8	71.2	300	11	C35496
C 50	17.8	71.2	300	11	C51717
C 51	17.8	71.2	300	11	C52955
C 52	17.8	71.2	316	13	FR0039786
C 53	17.8	71.2	342	10	AU184237
C 54	17.8	71.2	360	11	C67092
C 55	17.8	71.2	395	11	D27831
C 56	17.8	71.2	411	11	BF079024
C 57	17.8	71.2	436	10	BE025611
C 58	17.8	71.2	440	11	BE492205
C 59	17.8	71.2	449	10	AL584827
C 60	17.8	71.2	477	10	AU215151
C 61	17.8	71.2	540	11	BF491727
C 62	17.8	71.2	551	13	AO652018
C 63	17.8	71.2	579	13	AZ992973
C 64	17.8	71.2	591	13	AZ615095
C 65	17.8	71.2	598	13	FR0035666
C 66	17.8	71.2	600	13	FR0044579
C 67	17.8	71.2	611	13	FR0044632
C 68	17.8	71.2	612	10	AM638336
C 69	17.8	71.2	631	10	AU200354
C 70	17.8	71.2	655	10	AU202462
C 71	17.8	71.2	658	10	AU215033
C 72	17.8	71.2	664	13	TA243B05P
C 73	17.8	71.2	669	10	AI980907
C 74	17.8	71.2	680	13	AZ244185
C 75	17.8	71.2	706	10	AM821884
C 76	17.8	71.2	706	10	AM821888
C 77	17.8	71.2	772	10	AM773308
C 78	17.8	71.2	947	13	AZ135972
C 79	17.8	71.2	1015	13	CNS01050
C 80	17.6	70.4	150	11	BG265466
C 81	17.6	70.4	196	13	CNS03695
C 82	17.6	70.4	250	10	AM8756042
C 83	17.6	70.4	273	10	AM875988
C 84	17.6	70.4	285	10	AM876052
C 85	17.6	70.4	287	10	AA371107

AL124335 Fugu rubr
AL194151 Tetradon
AL230736 Tetradon
AI605313 vn87h03.x
AA624863 vn87h03.x
AI642276 vn87h03.y
BI141573 IPL_54.B0
AZ214122 Sheared D
BE990318 UI-M-B21
BG008569 PM4-CN030
BI141537 IPL_53.E1
AZ412964 IM0186F13
AI449930 mr81a05.x
AO522448 HS_5337.A
AL466271 T. brucei
AA527258 n935h02.s
AW702183 TGESt273
AZ244099 RPCI-23-3
BE656252 UI-M-BH0-
BE443324 WHE1112.E
AL455591 T. brucei
AZ003494 RPCI-23-3
AL472440 T. brucei
AO936593 HSJ41-840
AO941180 Sheared D
AM160263 EST230121
AL457556 T. brucei
AI325366 m129f01.y
AL322870 Tetradon
AK006021 Mus muscu
BE095496 UI-R-BU0-
BF428015 daa04908.
AL132206 Fugu rubr
AU074974 AU074974
BF440174 BS2900018
BB190815 BB190815
C35496 C35496 Yuj1
C51717 C51717 Yuj1
C52955 C52955 Yuj1
AL127281 Fugu rubr
AU184237 AU184237
C67092 C67092 Yuj1
D27831 CELK0045F
BF079024 229530.MA
BE025611 db25b07.Y
BF492205 AT29101.Y
AL584827 AL584827
AU215151 AU215151
BF491727 AT28535.5
AO652018 Sheared D
AZ992973 2M0277M14
AZ615095 IM044E16
AL123183 Fugu rubr
AL132071 Fugu rubr
AL131214 Fugu rubr
AM638336 b16bh04.w
AU200354 AU200354
AU202462 AU202462
AU215033 AU215033
AI481666 T. brucei
AI980907 pat.PR003
AZ244185 RPCI-23-1
AM821884 RCO-ST037
AM821888 RCO-ST037
AM773308 EST00214
AZ135972 SP_0170.A
AL069526 Drosophi1
BG265466 1000027B0
AL229753 Tetradon
AM876042 CM3-PT001
AM875988 CM3-PT001
AM876052 CM3-PT001
AA371107 EST82857

Thu Dec 27 16:12:12 2001

us-09-396-196f-3.rn1

Page 19

LOCATION: sequence of SEQ ID NO: 1;
IDENTIFICATION METHOD: Homology to MADS-box proteins
IDENTIFICATION METHOD: FEATURE:
NAME/KEY: K-Box
LOCATION: Residues 90-143 of deduced amino acid
LOCATION: sequence of SEQ ID NO: 1;
IDENTIFICATION METHOD: Homology to MADS-box proteins
US-08-485-981-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 tgcgtgacgcgtgttga 21
111 111 111111111
DB 495 TGCCTGATCAGCTGTTGA 513

RESULT 45
US-08-867-087B-1
Sequence 1, Application US/08867087B
Patent No. 5990386
GENERAL INFORMATION:
APPLICANT: An, Gynheung
TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT
TITLE OF INVENTION: AND APICAL DOMINANCE IN PLANTS
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klarquist Sparkman Campbell Leigh &
ADDRESSEE: Whinston, LLP
STREET: One World Trade Center
STREET: 121 S.W. Salmon Street
STREET: Suite 1600
CITY: Portland
STATE: Oregon
COUNTRY: United States of America
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: disk, 3-1/2 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,087B
FILING DATE: June 2, 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/323,449
FILING DATE: October 14, 1994
APPLICATION NUMBER: U.S. 08/485,981
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Alan E.
REGISTRATION NUMBER: 35,123
REFERENCE/DOCKET NUMBER: 4630-47071
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 226-7391
TELEFAX: (503) 228-9446
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1141 base pairs
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
US-08-867-087B-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 tgcgtgacgcgtgttga 21

DB 495 TGCCTGATCAGCTGTTGA 513
1111 111 111111111

Search completed: December 26, 2001, 12:23:04
Job time: 7141 sec

US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Paul J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 64.0%; Score 16; DB 4; Length 80161;
Best Local Similarity 79.2%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 gtgtgtgagatctgtctgttgaac 24
||||| 1 |||| 11 ||
DB 36246 gtgtgtgagagcggtgtgtgaagc 36269

RESULT 43
US-08-323-449B-1
; Sequence 1, Application US/08323449B
; Patent No. 5859326
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENE CONTROLLING FLORAL DEVELOPMENT AND
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klariquist Sparkman Campbell Leigh &
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,449B
; FILING DATE: October 14, 1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-41493
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141 base pairs

TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
FEATURE:
NAME/KEY: MADS box
LOCATION: Residues 2-57 of deduced amino acid
LOCATION: sequence of SEQ ID NO: 1;
IDENTIFICATION METHOD: Homology to MADS-box proteins
IDENTIFICATION METHOD: FEATURE:
NAME/KEY: K-box
LOCATION: Residues 90-143 of deduced amino acid
LOCATION: sequence of SEQ ID NO: 1;
IDENTIFICATION METHOD: Homology to MADS-box proteins
US-08-323-449B-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 tgcgtgagatctgtgttga 21
||||| 1 ||||| 11 ||
DB 495 TGCTTGATCAGCTGTGTGA 513

RESULT 44
US-08-485-981-1
; Sequence 1, Application US/08485981
; Patent No. 5861542
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENE CONTROLLING FLORAL DEVELOPMENT AND
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klariquist Sparkman Campbell Leigh &
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,981
; FILING DATE: June 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/323,449
; FILING DATE: October 14, 1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-42933
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; FEATURE:
NAME/KEY: MADS box
LOCATION: Residues 2-57 of deduced amino acid

Page 17

RESULT 42

Query Match 64.0%; Score 16; DB 3; Length 1892;
Best Local Similarity 79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ttgtcgtgactcgtcgtttgaagc 24
1115 GCTGTGATGTCAGCTTGTGAGC 1092

RESULT 37

US-08-676-967-3/C
Sequence 3, Application US/08676967
Patent No. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,967
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-676-967-3

Query Match 64.0%; Score 16; DB 1; Length 2733;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 ttgtcgtcgtcgtcgtttgaagc 25
2301 TTGTCACACGCTGCTTGAAGC 2278

RESULT 38

US-08-676-974-3/C
Sequence 3, Application US/08676974
Patent No. 5770422
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco

STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,974
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-676-974-3

Query Match 64.0%; Score 16; DB 1; Length 2733;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 ttgtcgtcgtcgtcgtttgaagc 25
2301 TTGTCACACGCTGCTTGAAGC 2278

RESULT 39

US-09-098-487-3/C
Sequence 3, Application US/09098487
Patent No. 5917025
GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2733 base pairs
TYPE: nucleic acid

```

FILE REFERENCE: CCDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO: 74
LENGTH: 599
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(599)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-74

```

```

Query Match          64.0%; Score 16; DB 4; Length 599;
Best Local Similarity 79.2%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 gtcgtgacgtcgtgttgaagc 24
    ||||| ||||| ||||| |||||
Db 367 gttcctgcattcttctgcctgaac 344

```

RESULT 35

```

US-08-933-750C-66/C
Sequence 66, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
FILING DATE:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

```

```

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT03
CLONE: 641127
US-08-933-750C-66

```

```

Query Match          64.0%; Score 16; DB 2; Length 1892;
Best Local Similarity 79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 gtcgtgacgtcgtgttgaagc 24
    ||||| ||||| ||||| |||||
Db 1115 gctggtgacgtgcagtttgagc 1092

```

RESULT 36

```

US-09-234-613-66/C
Sequence 66, Application US/09234613
Patent No. 6132973
GENERAL INFORMATION:

```

```

APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

```

```

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT03
CLONE: 641127
US-09-234-613-66

```

APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Tomato
FEATURE:
NAME/KEY: CDS
LOCATION: 194..3535
US-08-811-583-1

Query Match 65.6%, Score 16.4; DB 4; Length 3731;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ttgtgagatctgctgtt 18
|||||
Db 2422 GTTGCTGATCTGCTGTT 2405

RESULT 32
US-08-649-046-1
Sequence 1, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: OLSZEWSKI, NEIL E.
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETTING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3476 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-649-046-1

Query Match 64.8%; Score 16.2; DB 2; Length 3476;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtgagatctgctgttga 22
|||||
Db 2803 TGGCTGATCTCTGTTGTA 2823

RESULT 33
US-08-649-046-3
Sequence 3, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: OLSZEWSKI, NEIL E.
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETTING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-649-046-3

Query Match 64.8%; Score 16.2; DB 2; Length 6479;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtgagatctgctgttga 22
|||||
Db 5358 TGGCTGATCTCTGTTGTA 5378

RESULT 34
US-09-385-982-74/C
Sequence 74, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
PRODUCTS: II

```

; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3603 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2214
; US-08-188-582-15

Query Match
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttcgtgacatgctgttgaag 23
    ||||| ||||| |||||
Db 997 gctgctgctctgctgctgag 975

RESULT 29
US-08-646-715-15/c
; Sequence 15, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1889
; TELEFAX: (415) 398-3249
; TEXT: 910 277299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3603 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 1..2214
; US-08-646-715-15

Query Match
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttcgtgacatgctgttgaag 23
    ||||| ||||| |||||
Db 997 gctgctgctctgctgctgag 975

RESULT 30
US-09-058-489-13
; Sequence 13, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 5322
; TYPE: DNA
; ORGANISM: Human
; US-09-058-489-13

Query Match
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgacatgctgttgaag 24
    ||||| ||||| |||||
Db 2522 ttgctgacatgctgttgaagc 2544

RESULT 31
US-08-811-583-1/c
; Sequence 1, Application US/0881583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMACTIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRp)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```


Db 816 TTGCTGATCTGCTGTTTGAGC 838

RESULT 26
US-09-299-843A-31
Sequence 31, Application US/09299843A
Patent No. 6107475
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299, 843A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/088,337
FILING DATE: 01-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,848
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Jill E. Oul
REGISTRATION NUMBER: 43,213
REFERENCE/DOCKET NUMBER: 27866/32059B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
FEATURE:
NAME/KEY: CDS
LOCATION: 94..1158
US-09-299-843A-31

Query Match 66.4%; Score 16.6; DB 3; Length 3119;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 316 TTGCTGATCTGCTGTTTGAGC 338

RESULT 27
US-09-058-489-14
Sequence 14, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
EARLIER FILING DATE: 1998-04-10
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 3408
TYPE: DNA
ORGANISM: Human
US-09-058-489-14

Query Match 66.4%; Score 16.6; DB 3; Length 3408;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 2522 ttgtgagatctctgttggaac 2544

RESULT 28
US-08-188-582-15/c
Sequence 15, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/JAT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3836
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2254 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1657
US-08-153-848-27

Query Match 66.4%; Score 16.6; DB 1; Length 2254;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgagctgcgtgttgaagc 24
||| ||||| ||||| |||
Db 816 TTGCTGATCTGCTGTTGTAGC 838

RESULT 24
US-09-299-843A-27
Sequence 27, Application US/09299843A
Patent No. 6107475
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,843A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/088,337
FILING DATE: 01-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,848
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Jill E. Uhl
REGISTRATION NUMBER: 43,213
REFERENCE/DOCKET NUMBER: 27866/32059B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX:
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2254 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1657
US-09-299-843A-27

Query Match 66.4%; Score 16.6; DB 3; Length 2254;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgagctgcgtgttgaagc 24
||| ||||| ||||| |||
Db 816 TTGCTGATCTGCTGTTGTAGC 838

RESULT 25
PCT-US93-11153-27
Sequence 27, Application PC/TUS9311153
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: Novel Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11153
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2254 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1657
PCT-US93-11153-27

Query Match 66.4%; Score 16.6; DB 5; Length 2254;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgagctgcgtgttgaagc 24

```
; Sequence 1, Application US/08287001A
; Patent No. 5622861
; GENERAL INFORMATION:
; APPLICANT: KAPLAN, GERARDO
; APPLICANT: FEINSTONE, STEPHEN M.
; TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSER: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, The Candler Bldg, 127 Peachtree
; STREET: Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spiralt, Gwedolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..1551
; US-08-287-001A-1

Query Match      66.4%; Score 16.6; DB 1: Length 2093;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctgacatcgtctgttgaag 23
Db 1402 GTTGTTGAAATCCTCTTTTGAG 1380

RESULT 22
PCT-US95-09941-1/C
; Sequence 1, Application PC/TUS9509941
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSER: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, The Candler Bldg, 127 Peachtree
; STREET: Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,001A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/287,001
; FILING DATE: 5 AUG 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Spiralt, Gwedolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..1551
; PCT-US95-09941-1
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09941
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/287,001
FILING DATE: 5 AUG 1994
ATTORNEY/AGENT INFORMATION:
NAME: Spiralt, Gwedolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.621
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2093 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 196..1551
PCT-US95-09941-1

Query Match      66.4%; Score 16.6; DB 5: Length 2093;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctgacatcgtctgttgaag 23
Db 1402 GTTGTTGAAATCCTCTTTTGAG 1380

RESULT 23
US-08-153-848-27
; Sequence 27, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804e1 Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Marshall, O'Toole, Gerstein, Murray &
; STREET: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5759804e1, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
```


APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,991
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/94,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa-protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-08-876-991-1

Query Match 67.2% Score 16.8; DB 2; Length 12284;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
UY 6 tggatcgtcgtttgaagcg 25
|||||
LV 2249 TGGCTCTGCTGTTCAGAG 2268
RESULT 18
US-09-059-853-1
Sequence 1, Application US/09059853
Patent No. 5935582
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa-protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
US-09-059-853-1

```

STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,577
FILING DATE: 12-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,48554
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 43..3364
US-08-366-577-1

Query Match      68.0%; Score 17; DB 1; Length 3435;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgatactgtgttgaagcg 25
    |||||  |||||  |||||  |||||
Db 16 GTGGCGGGAACGCTGTTGAAGCG 40

RESULT 15
PCT-US96-00005-1
; Sequence 1, Application PC/TUS9600005
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL
; TITLE OF INVENTION: TUMORS WITH REPLICATION ERRORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 1001 G Street N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00005
; FILING DATE: 2-JAN-96
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.

```

```

REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,53505
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 43..3364
PCT-US96-00005-1

Query Match      68.0%; Score 17; DB 5; Length 3435;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgatactgtgttgaagcg 25
    |||||  |||||  |||||  |||||
Db 16 GTGGCGGGAACGCTGTTGAAGCG 40

RESULT 16
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patient No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Ashley, Gary
; APPLICANT: Jullien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020, 20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-144-085-3

Query Match      68.0%; Score 17; DB 4; Length 33529;
Best Local Similarity 80.0%; Pred. No. 1,1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgatactgtgttgaagcg 25
    |||||  |||||  |||||  |||||
Db 31369 gttgctgaagcggtgtctgaagcg 31393

RESULT 17
US-08-876-991-1
; Sequence 1, Application US/08876991
; Patient No. 5925360
; GENERAL INFORMATION:
; APPLICANT: Gregor Meyers, Tillmann R menapf.

```



```
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990, 823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 5
LENGTH: 456
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-5
```

```
Query Match          70.4%; Score 17.6; DB 4; Length 456;
Best Local Similarity 83.3%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 gtcgtgacatcgtctgttgaagc 24
      ||| ||||| ||||| || ||
Db      280 gttggtgacatcgtctgttgaagc 303
```

```
RESULT      8
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
```

```
Query Match          70.4%; Score 17.6; DB 4; Length 4403765;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 gtcgtgacatcgtctgttgaagc 24
      ||| ||||| ||||| || ||
Db      4096776 gttggtgacatcgtctgttgaagc 4096753
```

```
RESULT      9
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
```

```
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          70.4%; Score 17.6; DB 4; Length 4411529;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 gtcgtgacatcgtctgttgaagc 24
      ||| ||||| ||||| || ||
Db      4104579 gttggtgacatcgtctgttgaagc 4104556
```

```
RESULT      10
US-09-136-442-2/c
Sequence 2, Application US/09136442
Patent No. 6030825
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
TITLE OF INVENTION: CYCLOPHILIN-TYPE PEPTIDYL-PROLYL CIS/TRANS ISOMERASE
FILE REFERENCE: PF-0582 US
CURRENT APPLICATION NUMBER: US/09/136,442
CURRENT FILING DATE: 1998-08-19
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PERL Program
SEQ ID NO 2
LENGTH: 1069
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2925455
US-09-136-442-2
```

```
Query Match          68.8%; Score 17.2; DB 3; Length 1069;
Best Local Similarity 86.4%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      3 tgcgtgacatcgtctgttgaagc 24
      || ||||| ||||| || ||
Db      364 tcttgatctcctgttgaagc 343
```

```
RESULT      11
US-08-206-176-5/c
Sequence 5, Application US/08206176
Patent No. 5639940
GENERAL INFORMATION:
APPLICANT: Garner, Ian
APPLICANT: Dalrymple, Michael A
APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of Fibrinogen in Transgenic
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
```



```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/411,768B
: FILING DATE: 31-March-95
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: CH 3124/92
: FILING DATE: 02-OCT-1992
: PRIORITY APPLICATION NUMBER: CH 2134/93
: FILING DATE: 15-JUL-1993
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5872 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: STRAIN: DSM498
: IMMEDIATE SOURCE:
: CLONE: pBO30A15-9
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1154..2308
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /codon_start= 1154
: OTHER INFORMATION: /EC_number= 2.3.1.47
: OTHER INFORMATION: /product= "KAPA synthase"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
: OTHER INFORMATION: /gene= "biof"
: OTHER INFORMATION: /number= 2
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: OTHER INFORMATION: /EC_number= 6.3.3.3
: OTHER INFORMATION: /product= "DTP synthase"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
: OTHER INFORMATION: /gene= "biof"
: OTHER INFORMATION: /number= 4
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
: NAME/KEY: RBS
: LOCATION: 1141..1156
: OTHER INFORMATION: /standard_name= "biof RBS"
: FEATURE:
: NAME/KEY: RBS
: LOCATION: 3030..3045
: OTHER INFORMATION: /standard_name= "biof RBS"
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 87/01391 B1
: FILING DATE: 26-AUG-1986
: PUBLICATION DATE: 07-APR-1993
: US-08-411-768B-6
:
: Query Match
: Best Local Similarity 100.0%; Score 25; DB 3; Length 5872;
: Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1 gtctgcatctgtgttgaagc 25
: Db 173 GTTCTGATCTGCTGTTGAAGCG 197
:
: RESULT 5
: US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328

```

```

: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
: US-09-103-840A-2

```

```

: Query Match
: Best Local Similarity 72.8%; Score 18.2; DB 4; Length 4403765;
: Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

: QY 3 tgcgtgcatctgtgttgaagc 25
: Db 57414 tgcgtgcatctgtgttgaagc 57436

```

```

: RESULT 6
: US-09-103-840A-1
: Sequence 1, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv
: US-09-103-840A-1

```

```

: Query Match
: Best Local Similarity 72.8%; Score 18.2; DB 4; Length 4411529;
: Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

: QY 3 tgcgtgcatctgtgttgaagc 25
: Db 57472 tgcgtgcatctgtgttgaagc 57494

```

```

: RESULT 7
: US-08-990-823-5
: Sequence 5, Application US/08990823D
: Patent No. 6228371
: GENERAL INFORMATION:
: APPLICANT: Nano, Francis
: TITLE OF INVENTION: Mycobacterium tuberculosis DNA Sequences Encoding
: TITLE OF INVENTION: Immunostimulatory Peptides

```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biofin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioB"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="Involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioC"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /amino1transf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter placc"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter placc"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="bioB RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioC RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter placc"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtcgtgatactcgtcgttgaagcg 25
Db 173 GTTGTGATCTGCTGTTGAAGCG 197

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: wordperfect
SOFTWARE: Version 5.1

```

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match          100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtcttgaagcg 25
    |||||||||||||||||||
Db 57 GTTGCTGATCTGCTGTTGAAGCG 81

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match          100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtcttgaagcg 25
    |||||||||||||||||||
Db 57 GTTGCTGATCTGCTGTTGAAGCG 81

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:20:11 : Search time 141.92 Seconds
(without alignments)
39,895 Million cell updates/sec

Title: US-09-396-196F-3

Sequence: 1 gttgtgcatctgtctgttgaacg 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database:

Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	18.2	72.8	4403765	4	US-09-103-840A-2
6	18.2	72.8	4411529	4	US-09-103-840A-1
7	17.6	70.4	4403765	4	US-08-990-823-5
8	17.6	70.4	4403765	4	US-09-103-840A-2
9	17.6	70.4	4411529	4	US-09-103-840A-1
10	17.2	68.8	10659	2	US-09-136-442-2
11	17.2	68.8	10654	1	US-08-206-176-5
12	17.2	68.0	751	3	US-09-010-809-4
13	17.2	68.0	1008	1	US-08-252-966B-15
14	17.2	68.0	3435	5	PCT-US96-00005-1
15	17.2	68.0	3435	5	US-09-144-085-3
16	17.2	68.0	3435	5	US-08-876-991-1
17	17.2	68.0	3435	5	US-09-059-853-1
18	17.2	68.0	3435	5	US-08-153-848-31
19	17.2	68.0	3435	5	PCT-US93-11153-31
20	17.2	68.0	3435	5	US-08-287-001A-1
21	17.2	68.0	3435	5	PCT-US95-09941-1
22	17.2	68.0	3435	5	US-08-153-848-27
23	17.2	68.0	3435	5	US-09-299-843A-27
24	17.2	68.0	3435	5	PCT-US93-11153-27
25	17.2	68.0	3435	5	US-09-299-843A-31
26	17.2	68.0	3435	5	US-09-058-489-14
27	17.2	68.0	3435	5	Sequence 14, Appl

28	16.6	66.4	3603	1	US-08-188-582-15	Sequence 15, Appl
29	16.6	66.4	3603	1	US-08-646-715-15	Sequence 15, Appl
30	16.6	66.4	5322	3	US-09-058-489-13	Sequence 13, Appl
31	16.4	65.6	3731	4	US-08-811-583-1	Sequence 1, Appl
32	16.2	64.8	3476	2	US-08-649-046-1	Sequence 1, Appl
33	16.2	64.8	6479	2	US-08-649-046-3	Sequence 3, Appl
34	16.2	64.0	1892	4	US-09-385-982-74	Sequence 74, Appl
35	16.2	64.0	1892	4	US-08-933-750C-66	Sequence 66, Appl
36	16.2	64.0	1892	3	US-09-234-613-66	Sequence 66, Appl
37	16.2	64.0	1892	3	US-08-676-967-3	Sequence 3, Appl
38	16.2	64.0	2733	1	US-08-676-974-3	Sequence 3, Appl
39	16.2	64.0	2733	2	US-09-098-487-3	Sequence 3, Appl
40	16.2	64.0	5000	3	US-09-104-070-1	Sequence 1, Appl
41	16.2	64.0	80161	4	US-09-036-987A-1	Sequence 1, Appl
42	16.2	64.0	80161	4	US-09-370-700-1	Sequence 1, Appl
43	16.2	64.0	80161	4	US-08-323-449B-1	Sequence 1, Appl
44	15.8	63.2	1141	2	US-08-485-981-1	Sequence 1, Appl
45	15.8	63.2	1141	2	US-08-867-087B-1	Sequence 1, Appl
46	15.8	63.2	2032	4	US-09-045-284A-1	Sequence 1, Appl
47	15.8	63.2	2854	1	US-08-121-713D-57	Sequence 57, Appl
48	15.8	63.2	2854	1	US-08-835-268-57	Sequence 57, Appl
49	15.8	63.2	2854	2	US-09-060-692-57	Sequence 57, Appl
50	15.8	63.2	2854	3	US-08-833-391-57	Sequence 57, Appl
51	15.8	63.2	2854	5	PCT-US94-10151A-57	Sequence 57, Appl
52	15.8	63.2	3600	4	US-08-855-910-7	Sequence 1, Appl
53	15.8	63.2	28804	2	US-08-592-872-1	Sequence 1, Appl
54	15.8	63.2	28804	3	US-09-096-942-2	Sequence 2, Appl
55	15.8	63.2	28804	3	US-09-096-942-2	Sequence 2, Appl
56	15.6	62.4	491	4	US-09-006-867-2	Sequence 2, Appl
57	15.6	62.4	491	4	US-09-400-208B-23	Sequence 23, Appl
58	15.6	62.4	747	3	US-08-750-145A-19	Sequence 19, Appl
59	15.6	62.4	747	3	US-08-975-698A-23	Sequence 23, Appl
60	15.6	62.4	747	4	US-09-417-090-23	Sequence 23, Appl
61	15.6	62.4	1328	1	US-07-679-052A-14	Sequence 14, Appl
62	15.6	62.4	1340	1	US-07-679-052A-16	Sequence 16, Appl
63	15.6	62.4	1610	1	US-07-768-286B-1	Sequence 1, Appl
64	15.6	62.4	1920	1	US-08-487-753-1	Sequence 1, Appl
65	15.6	62.4	1920	2	US-08-480-065-1	Sequence 1, Appl
66	15.6	62.4	1920	5	US-08-487-744-1	Sequence 1, Appl
67	15.6	62.4	2230	3	PCT-US93-09167-1	Sequence 1, Appl
68	15.6	62.4	2367	1	US-08-844-188-34	Sequence 34, Appl
69	15.6	62.4	2409	1	US-08-441-139-3	Sequence 3, Appl
70	15.6	62.4	2409	1	US-07-911-531-18	Sequence 18, Appl
71	15.6	62.4	2441	1	US-07-693-636A-18	Sequence 18, Appl
72	15.6	62.4	3113	2	US-08-332-312-1	Sequence 2, Appl
73	15.6	62.4	3986	2	US-08-627-254C-27	Sequence 27, Appl
74	15.6	62.4	9432	2	US-08-227-231A-1	Sequence 1, Appl
75	15.6	62.4	9432	2	US-08-473-750-4	Sequence 4, Appl
76	15.6	62.4	9432	2	US-08-477-326-4	Sequence 4, Appl
77	15.6	62.4	9751	4	US-09-238-303-7	Sequence 7, Appl
78	15.6	62.4	30001	1	US-08-125-468-1	Sequence 1, Appl
79	15.6	62.4	30001	2	US-08-474-923-1	Sequence 3, Appl
80	15.4	61.6	26	2	US-08-199-722-3	Sequence 35, Appl
81	15.4	61.6	691	4	US-08-861-774E-35	Sequence 97, App
82	15.4	61.6	708	4	US-08-996-416-997	Sequence 6, Appl
83	15.4	61.6	714	2	US-08-768-964-6	Sequence 8, Appl
84	15.4	61.6	714	2	US-08-768-964-8	Sequence 8, Appl
85	15.4	61.6	714	3	US-09-005-299-6	Sequence 6, Appl
86	15.4	61.6	714	3	US-09-005-299-8	Sequence 8, Appl
87	15.4	61.6	714	4	US-09-515-431-6	Sequence 6, Appl
88	15.4	61.6	714	4	US-09-515-431-8	Sequence 8, Appl
89	15.4	61.6	719	2	US-08-983-409-6	Sequence 6, Appl
90	15.4	61.6	789	2	US-08-768-964-4	Sequence 4, Appl
91	15.4	61.6	789	2	US-08-768-964-5	Sequence 5, Appl
92	15.4	61.6	789	3	US-09-005-299-4	Sequence 4, Appl
93	15.4	61.6	789	3	US-09-005-299-5	Sequence 5, Appl
94	15.4	61.6	789	4	US-09-515-431-4	Sequence 4, Appl
95	15.4	61.6	789	4	US-09-515-431-5	Sequence 5, Appl
96	15.4	61.6	809	4	US-09-276-531-27	Sequence 27, Appl
97	15.4	61.6	1001	3	US-08-983-409-4	Sequence 4, Appl
98	15.4	61.6	1069	2	US-08-768-964-1	Sequence 3, Appl
99	15.4	61.6	1069	2	US-08-768-964-3	Sequence 3, Appl
100	15.4	61.6	1069	3	US-09-005-299-1	Sequence 1, Appl

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149923.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0158328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1574;
Best Local Similarity 80.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 gttgcgatactcgtgttgaacg 25
| ||||| ||||| |||||
Db 516 gatgcggatcgcggttaag 540

Search completed: December 26, 2001, 12:18:23
Job time: 6864 sec

PE 25-FEB-2000: 2000EP-0301439.
XX
PR 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
PR 29-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 23-APR-1999: 99US-0130891.
PR 28-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132487.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132863.
PR 11-MAY-1999: 99US-0134256.
PR 11-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136392.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142350.
PR 08-JUL-1999: 99US-0142803.

PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 03-AUG-1999: 99US-0146389.
PR 04-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.

PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1237;
 Best Local Similarity 80.0%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gtgtcgtgactgctgttgaagc 25
 | ||||| ||||| || |||||
 Db 235 gatgctggctcgtcgttgaag 259

RESULT 43

AA39643

ID AAX39643 standard; DNA: 1514 BP.

XX AAX39643;

XX 02-JUL-1999 (first entry)

XX Renal cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

XX Homo sapiens.

XX OS

XX PN

WO9904265-A2.

XX 28-JAN-1999.
 PD 15-JUL-1998; 98WO-US14679.
 XX 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX (LUDWIG) LUDWIG INST CANCER RES.
 XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
 PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 DR WPI: 1999-13248/11.
 XX
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 PS Claim 67; Page 434; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

SQ Sequence 1514 BP; 388 A; 398 C; 375 G; 349 T; 4 other;

Query Match 68.0%; Score 17; DB 20; Length 1514;
 Best Local Similarity 80.0%; Pred. No. 3.7e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gtgtcgtgactgctgttgaagc 25
 |||| ||||| ||||| || |||||
 Db 766 gtgtgagtgctcgtcgttgaagc 790

RESULT 44

AAC48922

ID AAC48922 standard; DNA: 1573 BP.

XX AAC48922;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 59265.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PN

06-SEP-2000.

XX PD

XX XX

Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.
Arabidopsis thaliana.

XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134378.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR	18-JUN-1999;	9905-0139462;
PR	18-JUN-1999;	9905-0139463;
PR	18-JUN-1999;	9905-0139750;
PR	18-JUN-1999;	9905-0139763;
PR	21-JUN-1999;	9905-0139817;
PR	22-JUN-1999;	9905-0139899;
PR	23-JUN-1999;	9905-0140353;
PR	23-JUN-1999;	9905-0140354;
PR	24-JUN-1999;	9905-0140635;
PR	24-JUN-1999;	9905-0140823;
PR	30-JUN-1999;	9905-0140991;
PR	30-JUN-1999;	9905-0141287;
PR	01-JUL-1999;	9905-0141842;
PR	01-JUL-1999;	9905-0142154;
PR	02-JUL-1999;	9905-0142305;
PR	06-JUL-1999;	9905-0142803;
PR	08-JUL-1999;	9905-0142920;
PR	12-JUL-1999;	9905-0142927;
PR	13-JUL-1999;	9905-01433542;
PR	14-JUL-1999;	9905-01433622;
PR	15-JUL-1999;	9905-0144005;
PR	16-JUL-1999;	9905-0144085;
PR	16-JUL-1999;	9905-0144086;
PR	19-JUL-1999;	9905-0144325;
PR	19-JUL-1999;	9905-0144331;
PR	19-JUL-1999;	9905-0144332;
PR	19-JUL-1999;	9905-0144333;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144332;
PR	20-JUL-1999;	9905-0144632;
PR	21-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-01448814;
PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145088;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145192;
PR	23-JUL-1999;	9905-0145145;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145224;
PR	26-JUL-1999;	9905-0145276;
PR	27-JUL-1999;	9905-0145913;
PR	27-JUL-1999;	9905-0145918;
PR	27-JUL-1999;	9905-0145919;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	03-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0146389;
PR	04-AUG-1999;	9905-0147038;
PR	04-AUG-1999;	9905-0147204;
PR	04-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147260;
PR	05-AUG-1999;	9905-0147260;
PR	06-AUG-1999;	9905-0147303;
PR	06-AUG-1999;	9905-0147416;
PR	06-AUG-1999;	9905-0147416;
PR	16-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149368;
PR	18-AUG-1999;	9905-0149175;
PR	20-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149722;
PR	20-AUG-1999;	9905-0149723;
PR	23-AUG-1999;	9905-0149929;
PR	23-AUG-1999;	9905-0149902;
PR	23-AUG-1999;	9905-0149930;

PR	25-AUG-1999;	9905-01505864
PR	26-AUG-1999;	9905-01505864
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-SEP-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151436
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0152366
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155139
PR	23-SEP-1999;	9905-0155486
PR	24-SEP-1999;	9905-0155659
PR	28-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0156596
PR	04-OCT-1999;	9905-0157117
PR	05-OCT-1999;	9905-0157753
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158620
PR	08-OCT-1999;	9905-0158623
PR	12-OCT-1999;	9905-01586369
PR	13-OCT-1999;	9905-0158293
PR	13-OCT-1999;	9905-0158294
PR	13-OCT-1999;	9905-01585295
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159330
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159638
PR	18-OCT-1999;	9905-0159584
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160814
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160982
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161360
PR	26-OCT-1999;	9905-0161361
PR	28-OCT-1999;	9905-0161922
PR	28-OCT-1999;	9905-0161922
PR	29-OCT-1999;	9905-01621493
PR	29-OCT-1999;	9905-01621493

```

Query Match: Score 17; DB 21; Length 1200;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0
OY 1 gttcctgatactcgtttgaagcg 25
    | | | | | | | | | | | | |
Db 200 gatgcctggtcgtcgtgttaagag 224

RESULT 42
AAC34154
ID AAC34154 standard; DNA; 1237 BP.
XX
AC AAC34154;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5643.
XX

```

ID	AAT70131/C
XX	AAT70131 standard; cDNA to mRNA: 1008 BP.
AC	AAT70131;
XX	
DT	07-FEB-1998 (first entry)
XX	
DE	Max-interacting protein coding sequence (clone 19).
XX	
KW	murine; msln: mammalian homologue; Saccharomyces cerevisiae; repressor; KX Sin3; Mad: Max; msln:Mad complex; mSin:Mad:Max complex; Myc: KW basic helix-loop-helix zipper protein; compete; DNA-binding; MYC:Myc complex; activate; transcription; gene regulation; ss. XX OS Mus musculus. XX PN US5624818-A. XX PD 29-APR-1997. XX PF 01-JUN-1994; 94US-0252966. XX PR 01-JUN-1994; 94US-0252966. PR 19-SEP-1991; 91US-0756195. PR 23-OCT-1992; 92US-0903710. PR 01-APR-1994; 94US-0222638. PA (HUTCH-) HUTCHINSON CANCER RES CENT FRED. PI Ayer DE, Eisenman RN. PX WPT: 1997-258216/23. DR XX PT msln nucleic acids encoding recombinant polypeptide(s) that PT associate with Mad polypeptide - are possible homologues of S. PT cerevisiae general repressor protein Sin3 XX PS Example 13; Fig 26; 11pp; English. XX CC This CDNA sequence, designated clone 19, encodes a novel murine basic CC helix-loop-helix zipper (bHLZip) protein which interacts with the CC bHLZip, Max. Max is an obligate partner for the DNA binding and CC transcriptional functions of Myc family proteins as well as for the Mad CC protein. Max is a stable, ubiquitously expressed protein which in general CC does not appear to be regulated during mitogenesis, the cell cycle, or CC differentiation. Other cDNA sequences encoding murine proteins (msln) CC which may be mammalian homologues of the Saccharomyces cerevisiae general CC repressor protein Sin3 are claimed. The msln proteins associate with Mad CC and Max to form a msln:Mad:Max complex which binds to a CAGCGT sequence CC in promoters. Mad:Max complexes repress, while Myc:Max complexes CC activate, transcription from promoters containing proximal CAGCGT binding CC sites for these proteins. Expression of Mad is closely linked to CC differentiation in at least two distinct cell lineages. The switch from CC Myc:Max to Mad:Max complexes may reflect the repression of transcription CC of Myc regulated genes by Mad. The DNA, vectors and host cells of the CC invention are useful for the recombinant production of msln proteins CC useful in elucidation of Mad repressor functions. XX SQ Sequence 1008 BP: 247 A; 295 C; 295 G; 171 T; 0 other:
Query Match	68.0%; Score 17; DB 18; Length 1008;
Best Local Similarity	80.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0	
OY	1 gtcgtgatactgcattgaagcg 25 DB 449 GCCTGAGGCCTGTCTTCCTGCG 425
RESULT 41	
ID AAC48910	
AAC48910 standard; DNA: 1200 BP.	

	XX	AC	AAC48910;	XX
	DT	XX	18-OCT-2000	(first entry)
	XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 59321.	
	XX	DE	Hybridisation assay; genetic mapping; gene expression control;	
	KW	KM	Protein identification; signal transduction pathway;	
	KM	KM	metabolic pathway; Promoter; termination sequence; ss.	
	OS	XX	Arabidopsis thaliana.	
	PN	XX	EPI033405-A2.	
	XX	XX	06-SEP-2000.	
	PD	XX	06-SEP-2000.	
	XX	PF	25-FEB-2000; Z000EP-0301439.	
	XX	PR	25-FEB-1999; 99US-0121825.	
	PR	PR	05-MAR-1999; 99US-0123180.	
	PR	PR	09-MAR-1999; 99US-0123548.	
	PR	PR	23-MAR-1999; 99US-0125788.	
	PR	PR	25-MAR-1999; 99US-0126264.	
	PR	PR	29-MAR-1999; 99US-0126785.	
	PR	PR	01-APR-1999; 99US-0127462.	
	PR	PR	06-APR-1999; 99US-0128234.	
	PR	PR	08-APR-1999; 99US-0128714.	
	PR	PR	16-APR-1999; 99US-0129845.	
	PR	PR	19-APR-1999; 99US-0130077.	
	PR	PR	21-APR-1999; 99US-0130449.	
	PR	PR	23-APR-1999; 99US-0130510.	
	PR	PR	23-APR-1999; 99US-0130891.	
	PR	PR	28-APR-1999; 99US-0131449.	
	PR	PR	30-APR-1999; 99US-0132048.	
	PR	PR	04-MAY-1999; 99US-0132407.	
	PR	PR	05-MAY-1999; 99US-0132484.	
	PR	PR	06-MAY-1999; 99US-0132485.	
	PR	PR	06-MAY-1999; 99US-0132486.	
	PR	PR	07-MAY-1999; 99US-0132487.	
	PR	PR	11-MAY-1999; 99US-0132863.	
	PR	PR	14-MAY-1999; 99US-0134218.	
	PR	PR	14-MAY-1999; 99US-0134219.	
	PR	PR	14-MAY-1999; 99US-0134221.	
	PR	PR	14-MAY-1999; 99US-0134370.	
	PR	PR	18-MAY-1999; 99US-0134768.	
	PR	PR	19-MAY-1999; 99US-0134941.	
	PR	PR	20-MAY-1999; 99US-0135124.	
	PR	PR	21-MAY-1999; 99US-0135353.	
	PR	PR	24-MAY-1999; 99US-0135629.	
	PR	PR	25-MAY-1999; 99US-0136021.	
	PR	PR	27-MAY-1999; 99US-0136392.	
	PR	PR	28-MAY-1999; 99US-0136782.	
	PR	PR	01-JUN-1999; 99US-0137222.	
	PR	PR	03-JUN-1999; 99US-0137528.	
	PR	PR	04-JUN-1999; 99US-0137502.	
	PR	PR	07-JUN-1999; 99US-0137724.	
	PR	PR	08-JUN-1999; 99US-0138094.	
	PR	PR	10-JUN-1999; 99US-0138540.	
	PR	PR	10-JUN-1999; 99US-0138847.	
	PR	PR	14-JUN-1999; 99US-0139119.	
	PR	PR	16-JUN-1999; 99US-0139452.	
	PR	PR	16-JUN-1999; 99US-0139453.	
	PR	PR	17-JUN-1999; 99US-0139457.	
	PR	PR	18-JUN-1999; 99US-0139454.	
	PR	PR	18-JUN-1999; 99US-0139455.	
	PR	PR	18-JUN-1999; 99US-0139456.	
	PR	PR	18-JUN-1999; 99US-0139457.	
	PR	PR	18-JUN-1999; 99US-0139458.	
	PR	PR	18-JUN-1999; 99US-0139459.	
	PR	PR	18-JUN-1999; 99US-0139460.	
	PR	PR	18-JUN-1999; 99US-0139461.	

PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 492;
 Best Local Similarity 80.0%; Pred. No. 3.2e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gtcgtgacatcgtctgttgaagc 25
 | ||||| ||||| || ||| |
 Db 242 gctgctggtcgtcgtgttaagag 266

RESULT 36
 AAV87783
 ID AAV87783 standard; cDNA; 729 BP.

AAV87783;

12-FEB-1999 (first entry)

EST clone ES306.

Expressed sequence tag; secreted protein; haematopoiesis regulator;
 tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

Homo sapiens.

MO9845437-A2.

15-OCT-1998.

10-APR-1998; 98MO-U506956.

10-APR-1997; 97US-0837312.

(GENY) GENETICS INST INC.

Agostino MJ, Jacobs K, Lavallic ER, McCoy JM, Merberg D;

Racie LA, Spaulding V, Treacy M;

WPI: 1999-070078/06.

New polynucleotides encoding human secreted proteins - derived from
 e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 ovary, pituitary, retina and colon cDNA libraries

Claim 1; Page 178; 64pp: English.

The present sequence represents an expressed sequence tag (EST), and is
 a polynucleotide of the invention. The polynucleotides of the invention
 are all secreted EST sequences isolated from a variety of human tissue
 sources. The EST sequences and proteins encoded by them are predicted to
 have useful biological activities which would make them suitable for
 treating, preventing or ameliorating medical conditions in humans and
 animals, although no supporting data is given. Suggested activities
 include nutritional activity, immune stimulating or suppressing activity,
 haematopoiesis regulating activity, tissue growth activity,
 activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 activity. The EST sequences are also stated to be useful for gene
 therapy.

Sequence 729 BP; 188 A; 188 C; 167 G; 186 T; 0 other;

Query Match 68.0%; Score 17; DB 20; Length 729;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gtcgtgacatcgtctgttgaagc 25
 | |||| | ||||| || |||||
 Db 532 gtcgaagtcctcgtcgttgaagc 556

RESULT 37
 AAA58914
 ID AAA58914 standard; DNA; 751 BP.

AAA58914;

20-OCT-2000 (first entry)

DNA encoding ketosynthase domain of epothilone polyketide synthase.

Epothilone polyketide synthase; PKS; ketosynthase; acyltransferase;
 acyl carrier protein; polyketide; ketoreductase; enoylreductase;
 dehydratase; epothilone; antibiotic; antitumor; ss.

Sorangium cellulosum.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 68.0%; Score 17; DB 21; Length 751;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Claim 6; Column 15-16; 39pp: English.

The present sequence encodes a ketosynthase domain of epothilone
 polyketide synthase (PKS). In type I or modular PKS enzymes, a set of
 separate catalytic active sites (each termed a domain. A set of which
 is termed a module) exists for each cycle of carbon chain elongation
 and modification. The minimal PKS module is typified by module 3 which
 contains a ketosynthase domain, an acyltransferase domain, and an acyl
 carrier protein domain. These three enzyme activities are sufficient to
 activate the 2-carbon extender unit and attach it to the growing
 polyketide molecule. Additional domain which may be present include a
 ketoreductase domain, an enoylreductase domain, and a dehydratase domain.
 The PKS polypeptide is useful for preparing epothilone in large
 quantities. Epothilone is a polyketide antibiotic that also has
 anticancer activity.

Sequence 751 BP; 103 A; 220 C; 302 G; 126 T; 0 other;

PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.

PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148177.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 30-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0158293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.

XX WP1: 2001-442253/47.
 DR P-PSDB: AAM39770.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1: SEQ ID NO 1129; 10078bp; English.
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 1317 BP; 381 A; 266 C; 281 G; 389 T; 0 other;
 SO
 Query Match 68.8%; Score 17.2; DB 22; Length 1317;
 Best Local Similarity 86.4%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 3 tgcgtgactgctgtttgaagc 24
 Db 616 TGTGTGATCTCCGTTTGAACC 595
 RESULT 30
 AAT03854/C
 ID AAT03854 standard; DNA: 10564 BP.
 XX
 AC AAT03854;
 XX
 DT 26-OCT-1996 (first entry)
 XX
 DE Human fibrinogen gamma chain coding sequence.
 XX
 KW Human fibrinogen gamma chain; transgenic animal milk; treatment;
 KW sheep beta-lactoglobulin signal peptide fusion protein;
 KW surgical adhesive; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH 5'UTR 1..1798
 FT /*tag= a
 FT 1799..1876
 FT /*tag= b
 FT /note= "1"
 FT 1877..1972
 FT /*tag= c
 FT /note= "1"
 FT 1973..2017
 FT /*tag= d
 FT /note= "2"
 FT 2018..2206
 FT /*tag= e
 FT /note= "2"
 FT 2207..2390
 FT /*tag= f
 FT /note= "3"
 FT 2391..2509
 FT Intron

FT /*tag= g
 FT /note= "3"
 FT 2510..2603
 FT /*tag= h
 FT /note= "4"
 FT 2604..4210
 FT /*tag= i
 FT /note= "4"
 FT 4211..4341
 FT /*tag= j
 FT /note= "5"
 FT 4342..4644
 FT /*tag= k
 FT /note= "5"
 FT 4645..4778
 FT /*tag= l
 FT /note= "6"
 FT 4779..5757
 FT /*tag= m
 FT /note= "6"
 FT 5758..5942
 FT /*tag= n
 FT /note= "7"
 FT 5943..7425
 FT /*tag= o
 FT /note= "7"
 FT 7426..7703
 FT /*tag= p
 FT /note= "8"
 FT 7704..9341
 FT /*tag= q
 FT /note= "8"
 FT 9342..9574
 FT /*tag= r
 FT /note= "9"
 FT 9575..10564
 FT /*tag= s
 FT 1799-9574
 FT /tag= t
 FT /note= "exons 1-9"
 PN W05523868-A1.
 XX
 PD 08-SEP-1995.
 XX
 PF 01-MAR-1995; 95WO-US02648.
 XX
 PR 03-MAR-1994; 94US-0206176.
 XX
 PA (PHAR-) PHARM PROTEINS LTD.
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Dairymple MA, Foster DC, Garner I, Prunkard DE;
 XX
 DR WP1: 1995-320582/41.
 DR P-PSDB: AAR82245.
 XX
 PT Production of fibrinogen in transgenic mammals - by introducing DNA
 PT segments into the germ line of a non-human mammal and collecting
 PT milk from female progeny.
 XX
 PS Disclosure; Page 57-67; 99pp; English.
 XX
 CC This sequence encodes the human fibrinogen gamma chain. It may
 CC be operably linked to the signal peptide, preferably of the sheep
 CC beta-lactoglobulin gene (see AAT03855) and, together with the
 CC fibrinogen A-alpha chain sequence (see AAT03853) and the B-beta
 CC chain sequence (see AAT03852), is introduced into a fertilised egg
 CC or the germ line of a non-human animal, preferably a sheep, pig,
 CC goat or cattle at a molar ratio of 0.5:1.0:0.5:1.0:0.5:1.0,
 CC respectively, for fusion protein gene expression in transgenic
 CC animal milk. The recombinantly produced active fibrinogen is
 CC useful in human and veterinary medicine, e.g. in the formulation

CC obtained the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

Sequence 434 BP; 136 A; 73 C; 101 G; 122 T; 2 other;

Query Match	Score	DB	Length
68.88;	17.2;	DB 21;	434;

Matches	19;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

Qy	3	tgcgtgac	ctgctgtt	gaagc	24
Db	297	tgTTGGATCTCCTGTTGAACC	276		

RESULT 26

ID	AAF13715	standard;	CDNA;	687	BP.

AC AAF13715;

DT 13-MAR-2001 (first entry)

DE ASpergillus oryzae EST SEQ ID NO:6238.

KM Multiple gene expression; filamentous fungal cell; ESF.
KM expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;
KM *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catabolic pathway engineering; ss.

OS *Aspergillus oryzae*.

PN W0200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK A.S.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPT; 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -

PS Claim 88; Page 2566; 3161pp; English.

CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway

CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AF074718 to AAP1247 represents ESTs from
CC *Fusarium venenatum*, AF112407 to AAP11853 represents ESTs from
CC *Aspergillus niger*, AF11854 to AP114878 represents ESTs from *Aspergillus*
CC *sp.*, AF114879 to AP115337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.

Sequence 687 BP; 185 A; 202 C; 166 G; 129 T; 5 other;

Query Match	Score	DB	Length
68.8%	17.2	21	687

Matches	19;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

```

OY      1  gttgtgatctgctgtttgaa 22
          || ||||| ||||| |||||
Db      565  GTCGCTGGAACTGCTGATTGAA 544

```

RESULT 27

ID AAZ95242 standard; DNA; 1068 BP.

AC AAZ95242;

DT 05-JUN-2000 (first entry)

DE Cyclophilin-type peptidyl prolyl cis/trans isomerase nucleotide sequence.

KM Cyclophilin-type peptidyl prolylase/trans isomerase: CPCI; cancer; AIDS
 KM leukemia; reproductive disorder; asthma; diabetes; infertility; anaemia
 KM polycystic ovary syndrome; uterine fibroid; Good pasture's syndrome;
 KM gout; Grave's disease; multiple sclerosis; lupus; osteoarthritis;
 KM irritable bowel syndrome; ds.

OS Homo sapiens.

US6030825-A. PN

PD 29-FEB-2000.

PF 19-AUG-1998; 98US-0136442

PR 19-AUG-1998; 98US-0136442

PA (INCY-) INCYTE PHARM INC.

PI Hillman JL, Corley NC, Patterson C, Guegler KJ;

DR WPT; 2000-205207/18.

XX

PT Isolated polynucleotides encoding cyclophilin-type peptidyl-prolyl
cis/isomerase, useful for preventing, diagnosing and treating
M. tuberculosis/HIV-related lung disorders and reproductive diseases -

PS Claim 1; Fig 1; 28pp; English.

This sequence represents a human cyclophilin-type peptidyl-prolyl cis/trans isomerase (CPCI) nucleotide sequence. The invention includes probes for the CPCI nucleotide sequence and vectors expressing the polynucleotide. CPCI is a member of the peptidyl-prolyl cis/trans isomerase (PPIase) class of enzymes. Cyclophilin isomerase activity is essential for correct protein folding and protein trafficking. The CPCI nucleotide sequence and the protein it encodes may be used in the diagnosis, prevention and treatment of disorders associated with inappropriate CPCI expression and activity. For example, they may be used to treat cancers (e.g. leukaemia, lymphoma, melanoma and cancers of the breast, liver and prostate), autoimmune/inflammatory disorders (e.g. AIDS, asthma and diabetes mellitus) and reproductive disorders

and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impaired respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 8631 BP; 2280 A; 2050 C; 2270 G; 2031 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 8631;

Best Local Similarity 94.7%; Pred. No. 3.1e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 7591 TGCTGGAGCTGCTGTTGA 7573

RESULT 24

AAA35028/c

AAA35028;

28-JUL-2000 (first entry)

Human adenosine receptor related polynucleotide SEQ ID NO:2117.

Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antisthmatic; cytostatic; analgesic; impaired airway; lung disease; ischemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukemia; lymphoma; carcinoma; metastasis; ss.

Homo sapiens.

MO200009525-A2.

24-FEB-2000.

03-AUG-1999; 99MO-US17712.

03-AUG-1998; 98US-0095212.

(UYEC-) UNIV EAST CAROLINA.

Nyce JW;

WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers

Disclosure; Page 969-971; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antisthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischemic conditions, pulmonary vasoconstriction, allergies,

asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 185, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 8631 BP; 2280 A; 2050 C; 2270 G; 2031 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 8631;

Best Local Similarity 94.7%; Pred. No. 3.1e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 7591 TGCTGGAGCTGCTGTTGA 7573

RESULT 25

AAC03560/c

AAC03560;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 3558.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

P-PSDB; AAG03554.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures

Claim 1; SEQ ID 3558; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from cDNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been

XX Sequence 2009 BP; 418 A; 587 C; 643 G; 356 T; 5 other;

Query Match 69.6%; Score 17.4; DB 20; Length 2009;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tgcctgacgtcgtttga 21
||||||| |||||||||
Db 959 TGCTGGAGCTGCTGTGA 941

RESULT 22
AAZ07514/C
ID AAZ07514 standard; DNA: 2034 BP.

XX AAZ07514;
XX 26-NOV-1999 (first entry)

DE Human RIP-associated protein (RAP-2) clone #41072 nucleotide sequence.

KW Receptor interacting protein; RIP-associated protein-2; RAP-2; RIP;
KW inflammation; cell death; cell survival; septic shock; hepatitis;
KW graft versus host rejection; diabetes; multiple sclerosis; tumor;
KW HIV infection; p55-receptor; FAS-receptor; human; ss.

XX Homo sapiens.

OS Homo sapiens.

FN WO9947672-A1.

XX 23-SEP-1999.

PF 18-MAR-1999; 99WO-IL00158.

XX 19-MAR-1998; 98IL-0123758.

PR 01-SEP-1998; 98IL-0126024.

XX (YEDA) YEDA RES & DEV CO LTD.

PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

XX Wallach D, Kovalenko A, Horwitz MS, Li Y;

PI WPI: 1999-562113/47.

DR New receptor interacting protein-associated protein-2, used to develop

XX products for treating, e.g. septic shock, tumors or HIV infection

PT Claim 5; Fig 2A-B; 132pp; English.

XX The invention relates to areceptor interacting protein (RIP)-associated

CC protein-2 (RAP-2). The RAP-2 proteins, isoforms, analogs, fragments or

CC derivatives or DNA can be used for the modulation or mediation of the

CC RIP modulated/intracellular effects on the inflammation, cell

CC death or cell survival pathways in which RIP is involved directly, or

CC indirectly via other modulators/mediators of these pathways. They can be

CC used for treating e.g. septic shock, graft versus host rejection, acute

CC hepatitis, diabetes or multiple sclerosis. They can also be used for

CC treating tumor cells or HIV-infected cells or other diseased cells. The

CC RAP-2 binding proteins can also be used for modulating/mediating the function

CC of RAP-2. The products can also be used for diagnostic purposes, e.g.

CC for identifying disorders related to abnormal functioning of cellular

CC effects mediated by the p55-R, FAS-R or other related receptors. The

CC present sequence represents the nucleotide sequence of RAP-2 clone

CC #41072. This sequence was identical to a 1.5 Kb clone (AAZ07513) in its

CC coding region, but showed differences in the 5' non-coding region. Both

CC these sequences are assumed to be alternatively spliced forms of the same

XX Sequence 2034 BP; 427 A; 588 C; 633 G; 382 T; 4 other;

Query Match 69.6%; Score 17.4; DB 20; Length 2034;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tgcctgacgtcgtttga 21
||||||| |||||||||
Db 983 TGCTGGAGCTGCTGTGA 965

RESULT 23
AAF21150/C
ID AAF21150 standard; DNA: 8631 BP.

XX AAF21150;

XX 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2717.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;

KW human; airway disorder; bronchoconstriction; lung inflammation;

KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;

KW immunosuppressive; antisthmatic; analgesic; hypotensive; cytosolic;

KW respiratory obstruction; pulmonary obstruction; impeded respiration;

KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;

KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;

KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;

KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;

KW cancer; ss.

XX Homo sapiens.

OS WO200062736-A2.

XX 26-OCT-2000.

PF 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

PR (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

XX Nyce JW;

PI WPI: 2000-679539/66.

DR Low adenosine (A) content antisense oligonucleotides which do not

XX trigger adenosine receptors during metabolism, useful e.g. for treating

PT cancers and respiratory obstructions -

XX Disclosure; Page 1046-1048; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense

CC oligonucleotides and compositions (I) comprising them. In the antisense

CC oligonucleotides the A is replaced by a 'universal' or alternative base.

CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,

CC immunosuppressive, antisthmatic, hypotensive and cytostatic activities.

CC The antisense oligonucleotides and (I) can be used to down-regulate the

CC expression and/or activity of target polypeptides associated with

CC lung/respiratory disorders and malignancies, such as stimulating and

CC activating peptide factors and transmitters, transcription factors,

CC immunoglobulins and antibodies, antibody receptors, cytokines and

CC chemokines, endogenously produced specific and non-specific enzymes,

CC binding proteins, adhesion molecules and their receptors, cytokine and

CC chemokine receptors, adenosine receptors, bradykinin receptors, central

CC nervous system (CNS) and peripheral nervous and non-nervous system peptide

CC receptors, CNS and peripheral nervous and non-nervous system peptide

CC transmitters, defensins, growth factors, vasoactive peptides and

CC receptors, binding proteins and malignancy associated proteins. The

CC antisense oligonucleotides may be used in this way to treat disorders

CC including respiratory obstruction (especially pulmonary obstruction

CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)

CC The present sequence represents a human IKK-gamma subunit cDNA.
XX
SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;
Query Match 69.6%; Score 17.4; DB 21; Length 1994;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tgcgtgactcgtcttga 21
||||| |||||
Db 954 TGCTGAGAGCTGCTTTGA 936
RESULT 20
AAA5027/c
ID AAA5027 standard; DNA: 1994 BP.
XX
AC AAA5027;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2716.
XX
KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PE 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI: 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 968-969; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC sarcomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA52313 to AAA53312 represent the

CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA52323 to AAA53992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;
Query Match 69.6%; Score 17.4; DB 21; Length 1994;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tgcgtgactcgtcttga 21
||||| |||||
Db 954 TGCTGAGAGCTGCTTTGA 936
RESULT 21
AAZ07513/c
ID AAZ07513 standard; DNA: 2009 BP.
XX
AC AAZ07513;
XX
DT 26-NOV-1999 (first entry)
XX
DE Human RIP-associated protein (RAP-2) encoding DNA.
XX
KW Receptor interacting protein; RIP-associated protein-2; RAP-2; RIP;
KW inflammation; cell death; cell survival; septic shock; hepatitis;
KW graft versus host rejection; diabetes; multiple sclerosis; tumor;
KW HIV infection; p53-receptor; FAS-receptor; human; ss.
XX
OS Homo sapiens.
XX
PN WO947672-A1.
XX
PD 23-SEP-1999.
XX
PE 18-MAR-1999; 99WO-IL00158.
XX
PR 19-MAR-1998; 98IL-0123758.
XX
PR 01-SEP-1998; 98IL-0126024.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Wallach D, Kovalenko A, Horwitz MS, Li Y;
XX
DR WPI: 1999-562113/47.
XX
P-PSDB: AAY27430.
XX
PT New receptor interacting protein-associated protein-2, used to develop
PT products for treating, e.g. septic shock, tumors or HIV infection -
XX
PS Claim 4; Fig 1A-B; 132pp; English.
XX
CC This DNA encodes a receptor interacting protein (RIP)-associated protein
CC -2 (RAP-2). The RAP-2 proteins, isoforms, analogs, fragments or
CC derivatives or DNA can be used for the modulation or mediation of the
CC RIP modulated/mediated intracellular effects on the inflammation, cell
CC death or cell survival pathways in which RIP is involved directly, or
CC indirectly via other modulators/mediators of these pathways. They can be
CC used for treating e.g. septic shock, graft versus host rejection, acute
CC hepatitis, diabetes or multiple sclerosis. They can also be used for
CC treating tumor cells or HIV-infected cells or other diseased cells. The
CC RAP-2 binding proteins can be used for modulating/mediating the function
CC of RAP-2. The products can also be used for diagnostic purposes, e.g. for
CC identifying disorders related to abnormal functioning of cellular effects
CC mediated by the p55-R, FAS-R or other related receptors.

CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 Other;
SQ

Query Match 70.4%; Score 17.6; DB 21; Length 1437668;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 tctgtagctgtctgttgaagc 25
||||| ||||| ||||| ||||| |||||
Db 184667 ttgagatctgcggttgaagc 184690

RESULT 18
AAF21149/c
ID AAF21149 standard; DNA: 1994 BP.

XX AAF21149;
AC XX
DT 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2716.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; PDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.

XX Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

XX (NYCE/) NYCE J W.

XX Nyce JW;

XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -

XX Disclosure: Page 1045-1046; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (1) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (1) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central

CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (PDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

XX Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tctgtagctgtgttga 21
||||| ||||| ||||| ||||| |||||
Db 954 tctgtagctgtgttga 936

RESULT 19
AAC81426/c
ID AAC81426 standard; cDNA: 1994 BP.

XX AAC81426;
AC XX
DT 23-FEB-2001 (first entry)

XX Human I-kappa-B kinase gamma-subunit (IKK-gamma) cDNA.

XX Human; I-kappa-B kinase; IKK; antisense therapy; gene therapy;
KW cytokine expression inhibition; NF-kappa-B activation inhibition;
KW nuclear factor-kappa-B; Rheumatoid arthritis; immune disorder;
KW cancer; IKK-gamma; gamma-subunit; ss.

XX Homo sapiens.

XX JP2000253884-A.

XX 19-SEP-2000.

XX 10-MAR-1999; 99JP-0063291.

XX 10-MAR-1999; 99JP-0063291.

XX (TOAG) TOA GOSSEI CHEM IND LTD.

XX WPI; 2000-658813/64.

XX Antisense nucleic acid compound complementary to the subunit of
PT Ikappab, used to treat rheumatic arthritis, immune diseases and cancer

XX Claim 3; Page 14-15; 20pp; Japanese.

XX The invention relates to an antisense oligonucleotide targeted to
CC a gene encoding a subunit of I-kappa-B kinase (IKK) which inhibits its
CC expression, and thereby inhibits expression of a cytokine such as
CC IL-6 (interleukin-6). I-kappa-B kinase activates NF-kappa B (nuclear
CC factor-kappa-B) which acts as a transcriptional regulator of cytokine
CC genes. The antisense oligonucleotide can be used in gene therapy to
CC treat rheumatoid arthritis, immune disorders and cancers. Sequences
CC AAC81422-C81426 are cDNAs derived from genes whose expression may be
CC inhibited using an antisense oligonucleotide of the invention.

CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

XX Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 349980;

Best Local Similarity 83.3%; Pred. No. 3.7e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgtgactcgtcgttgaagc 25

Db 184668 ttgcagaatcgtccgttgaagc 184691

RESULT 16

AAH64966/C

ID AAH64966 standard; DNA: 349980 BP.

XX

XX AAH64966;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 1.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EPI108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOCYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX

XX Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.

XX Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 349980;

Best Local Similarity 83.3%; Pred. No. 3.7e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gttcgtgactcgtcgttgaagc 24

Db 23388 GCTGCTGATCTGTTGATGACG 23365

RESULT 17

AAAB1490

ID AAAB1490 standard; DNA: 1437668 BP.

XX AAAB1490;

XX 04-DEC-2000 (first entry)

XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

XX antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHTR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tetrelm H, Venter JC;

XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;

XX Rappelli R, Pizzo M;

XX WPI: 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be

XX used in the diagnosis and treatment of N. meningitidis infection and

XX other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAAB1254 to
XX AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to
XX AAAB1452 represent Neisseria meningitidis MenB polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a
XX medicament) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious
XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX Multivalent vaccines have also been tried but none have successfully
XX overcome antigenic variability. The provision of further, complete
XX sequences may provide an opportunity to identify secreted or surface

CC caused by bacteria.
 XX
 SO Sequence 4610 BP; 1106 A; 1083 C; 1118 G; 1303 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 4610;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gtgtgagatctgtgttgaagc 24
 |||||
 Db 4015 gtgtgagatctgtgttgaagc 4038

RESULT 14
 AAA81478/c
 ID AAA81478 standard; DNA: 52253 BP.

AC AAA81478;
 XX
 DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm_26 SPQ ID NO:26.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KM Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappunli R, Pizza M;

XX WPI: 2000-318079/27.

PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 532-547; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25665 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81234 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully

CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;

Query Match 70.4%; Score 17.6; DB 21; Length 52253;
 Best Local Similarity 83.3%; Pred. No. 3.1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtgagatctgtgttgaagc 25
 |||||
 Db 43436 ttgcacacatctgtgcgttgacgcg 43413

RESULT 15
 AAF21544
 ID AAF21544 standard; DNA: 345980 BP.

AC AAF21544;

XX 13-MAR-2001 (first entry)

DE Neisseria meningitidis B nucleotide sequence SPQ ID NO:1.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KM ds.

XX Neisseria meningitidis.

XX WO2000066791-A1.

XX 09-NOV-2000.

XX 08-MAR-2000; 2000WO-US05928.

XX 30-APR-1999; 99US-0132068.

XX 08-OCT-1999; 99WO-US23573.

XX 28-FEB-2000; 2000GB-0004695.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappunli R;
 PI Frazer CM, Grandi G;

XX WPI: 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -
 XX

XX Claim 7; Appendix A: 692pp; English.

XX The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the

CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

XX
SQ Sequence 3043 BP: 819 A; 624 C; 668 G; 928 T; 4 other;

Query Match 70.4%; Score 17.6; DB 21; Length 3043;

Best Local Similarity 83.3%; Pred. No. 2.2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gttgtgagatcgtctgttgaagc 24

Db 2421 gttgtgattgtctgttcaagc 2444

RESULT 12

AA33946 standard; DNA: 4573 BP.

XX
AC AAX33946;

DT 30-JUN-1999 (first entry)

DE Human HCMV inducible gene, SEQ ID NO 19.

KW HCMV inducible gene; cig; human; human cytomegalovirus; interferon;

KM anti-viral therapy; anti-HCMV therapy; detection; diagnosis;

OS Homo sapiens.

PN WO9913075-A2.

XX
PD 18-MAR-1999.

PF 08-SEP-1998; 98WO-US18638.

PR 22-SEP-1997; 97US-0059725.

PR 08-SEP-1997; 97US-0058180.

XX
PA (UYPR-) UNIV PRINCETON.

XX
PI Cong J, Schenk T, Zhu H;

XX
DR WPI: 1999-243729/20.

XX
DR P-PSDB; AAY05376.

XX
PT New isolated human genes

XX
PS Claim 2; Page 140-142; 184pp; English.

CC This sequence represents a human gene of the invention, that is induced
CC to express by both HCMV and interferon (IFN), designated HCMV-inducible
CC genes (cig or cigs). The invention also relates to genes that are
CC repressed in the presence of HCMV infection, designated HCMV-repressible
CC genes (crg or crgs). The products can be used to obtain agents which can
CC be used for anti-viral therapy, particularly anti-HCMV therapy. They can
CC also be used for the development of drugs that would allow for higher
CC dosage IFN treatments without the concomitant toxicity normally
CC associated with administering high levels of IFN. The products can also
CC be used for detection, diagnosis and drug screening.

XX
SQ Sequence 4573 BP: 1098 A; 1076 C; 1098 G; 1301 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 4573;

Best Local Similarity 83.3%; Pred. No. 2.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gttgtgagatcgtctgttgaagc 24

Db 3983 gttgtgaattgtctgttcaagc 4006

RESULT 13

AAF59633 standard; cDNA: 4610 BP.

XX
AC AAF59633;

DT 24-APR-2001 (first entry)

DE Human cell cycle and proliferation protein CCYPR-44 cDNA, SEQ ID NO:98.

XX
KW Cell cycle and proliferation protein; CCYPR; human; agonist;

KW antagonist; gene therapy; detection; gene therapy;

KW transgenic animal disease model; immune disorder;

KW developmental disorder; cell signalling disorder;

KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

KW arteriosclerosis; asthma; allergy; diabetes mellitus;

XX
KW menstrual cycle disorder; bacterial infection; ss.

OS Homo sapiens.

PN WO200107471-A2.

XX
PD 01-FEB-2001.

PF 21-JUL-2000; 2000WO-US19948.

PR 21-JUL-1999; 99US-0145075.

PR 08-SEP-1999; 99US-0153129.

PR 10-NOV-1999; 99US-0164647.

XX
PA (INCY-) INCYTE GENOMICS INC.

XX
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

XX
PI Azimzal Y, Yang J, Lu DM, Baughn MR, Patterson C, Shah P;

XX
DR WPI: 2001-112727/12.

XX
DR P-PSDB: AAB60496.

XX
PT Human cell cycle and proliferation proteins and polynucleotides are

XX
PT used to treat, diagnose and prevent immune, developmental and cell

XX
PT signalling disorders and cell proliferative disorders including cancer -

XX
PS Claim 5; Page 197-198; 205pp; English.

CC Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections

PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032136.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032232.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 PA (BADI) BASF AG.
 XX
 XX
 PI Pompeius M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 XX
 DR WPI; 2001-137957/14.
 DR P-PSDB; AAB80176.
 XX
 PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX
 XX Claim 3; Page 1637-1639; 1737pp; English.
 CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum

CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carboxylates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 SQ Sequence 1359 BP; 345 A; 394 C; 322 G; 298 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 22; Length 1359;
 Best Local Similarity 83.3%; Pred. No. 2e+02; Mismatches 0;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 DY 1 gttgcgagactcgtctgttgaagc 24
 1 |||||
 Db 1325 GCTGCTGATCTGTGAATGAAGC 1302
 RESULT 11
 AAC78167
 ID AAC78167 standard; cDNA; 3043 BP.
 XX
 AC AAC78167;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 XX Human cancer associated gene sequence SEQ ID NO:561.
 DE
 XX Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerrary; immunomodulator;
 KW antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;
 KW antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.
 KW
 XX Homo sapiens.
 OS
 XX Wo200055350-A1.
 PN
 XX 21-SEP-2000.
 PD
 XX 08-MAR-2000; 2000WO-US05882.
 PF
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI WPI: 2000-587533/55.
 DR P-PSDB; AAB43958.
 DR
 XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 PT
 XX Claim 1; Page 1086-1087; 2352pp; English.
 XX
 XX AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnerrary; immunomodulator;
 CC antidiabetic; antiaesthetic; antirheumatic; antiarthritic;
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating

PD 11-NOV-1999. 99WO-US09346.
 XX 30-APR-1999: 98US-0083758.
 XX 01-MAY-1998: 98US-0094869.
 XX 31-JUL-1998: 98US-0098994.
 XX 02-SEP-1998: 98US-0099062.
 XX 09-OCT-1998: 98US-0103749.
 XX 09-OCT-1998: 98US-0103794.
 XX 09-OCT-1998: 98US-0103796.
 XX 25-FEB-1999: 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masiangni V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 DR P-PSDB: AAY75109.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 7: Page 872; 1453pp: English.
 PS AA53015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medications for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX SO Sequence 774 BP: 137 A; 206 C; 210 G; 221 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 774;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgactctgctgttgaagcg 25
 ||||| ||||| ||||| ||||| |||||
 Db 57 ttgcagaatctgcgcgttgaagcg 80

RESULT 9
 AAH64991/c
 ID AAH64991 standard; DNA: 1236 BP.
 XX AAH64991:
 AC 26-SEP-2001 (first entry)
 DT C glutamicum coding sequence fragment SEQ ID NO: 26.
 DE C glutamicum coding sequence fragment SEQ ID NO: 26.
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX Corynebacterium glutamicum.
 OS EP108790-A2.
 PN 20-JUN-2001.
 PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999: 99JP-0377484.
 XX 07-APR-2000; 2000JP-0159162.
 XX 03-AUG-2000; 2000JP-0280988.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI: 2001-376931/40.
 DR P-PSDB: AAG89772.
 PT Novel polynucleotides derived from *Coryneform* bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX Claim 8: SEQ ID NO: 26; 246pp + Sequence Listing; English.
 PS The present invention provides a number of nucleotide and protein
 CC sequences from the *Coryneform* bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of *Coryneform* bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from *Coryneform* bacterium, and identifying a homologue of a gene derived
 CC from *Coryneform* bacterium. *Coryneform* bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 XX European Patent Office.
 XX SO Sequence 1236 BP: 306 A; 363 C; 299 G; 268 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 1236;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctgactctgctgttgaagc 24
 ||||| ||||| ||||| ||||| |||||
 Db 1225 GCTGCTGATCTGTTCATGATGAAGC 1202

RESULT 10
 AAF72295/c
 ID AAF72295 standard; DNA: 1359 BP.
 XX AAF72295:
 AC 30-APR-2001 (first entry)
 DT Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:1085.
 DE Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 XX fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme; ds.
 XX Corynebacterium glutamicum.
 OS WO200100843-A2.
 PN 04-JAN-2001.
 XX 23-JUN-2000; 2000WO-1B00923.
 XX 25-JUN-1999; 99US-0141031.
 XX 01-JUL-1999; 99DE-1030476.

```

XX
XX 23-SEP-1997 (first entry)
DE Partial DNA clone Acit#1-426 encoding immunostimulatory peptide.
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX W09700067-A1.
PN
XX 03-JAN-1997.
PD
XX 14-JUN-1996; 96WO-US10375.
PF
XX 15-JUN-1995; 95US-0000254.
PR
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
XX
XX WPI; 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 34; 79pp: English.
PS
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
CC encoded by this sequence has amino acid similarity to a dipeptide
CC transport protein.
XX
XX Sequence 456 BP; 63 A; 121 C; 165 G; 98 T; 9 other;
SQ

```

```

Query Match 70.4%; Score 17.6; DB 18; Length 456;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gtgtcgtgactgctgttgaagc 24
    ||||| ||||| ||||| || ||
Db 280 gtgtggtgactgctgtatgcgc 303

```

```

RESULT 7
AAC55976/c
ID AAC55976 standard; DNA: 527 BP.
XX
XX AAC55976;
AC
XX
XX 25-JAN-2001 (first entry)
DT
XX
XX Eucalyptus grandis transcription factor DNA sequence #107.
DE
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KM

```

```

KM poplar; sweetgum; teak; mahogany; bzrp; G-box binding factor;
KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KM homeodomain zipper; LIM domain; AP2; ERFBs; zinc finger domain;
KM type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX Eucalyptus grandis.
OS
XX W0200053724-A2.
PN
XX
XX 14-SEP-2000.
PD
XX
XX 09-MAR-2000; 2000WO-US06112.
PF
XX 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESTIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
PI
XX
XX WPI; 2000-579369/54.
DR
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
XX
XX Claim 1; Page 72; 747pp: English.
PS
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bzrp, bzrp family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERFBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
XX Sequence 527 BP; 97 A; 166 C; 173 G; 91 T; 0 other;
SQ

```

```

Query Match 70.4%; Score 17.6; DB 21; Length 527;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gtgtcgtgactgctgttgaagc 24
    ||||| ||||| ||||| |||||
Db 485 GCTGCTGCACGCTCTTTGAAGC 462

```

```

RESULT 8
AA253871
ID AA253871 standard; DNA: 774 BP.
XX
XX AA253871;
AC
XX
XX 21-MAR-2000 (first entry)
DT
XX
XX Neisseria meningitidis ORF 568 partial DNA sequence SEQ ID NO:1691.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy; ds.
XX
XX Neisseria meningitidis.
OS
XX
XX W09957280-A2.
PN
XX

```

```

AC  AA062386;
XX
XX  16-NOV-1994 (first entry)
XX
DE  Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW  Biotin; expression: enterobacteria; vitamin H; synthesis:
KW  plasmid: pB030A-15/9; bioB; bioC; bioD; bioA;
KW  promoter plac: biotin synthase; KAPA synthase;
KW  8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;
KW  dehydrobiotin synthase; DAPA synthase;
KW  S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW  seborrhoea; dermatitis; ds.
XX
XX  Escherichia coli DSM498.
XX
FH  Key
FT  Location/Qualifiers
FT  1..96
FT  /tag= a
FT  /function= "promoter plac"
FT  /evidence= EXPERIMENTAL
FT  23..28
FT  /tag= b
FT  /standard_name= "promoter plac"
FT  45..50
FT  /tag= c
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "promoter plac"
FT  105..109
FT  /tag= d
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "bioB RBS no. 9"
FT  117..1157
FT  /tag= e
FT  /product= "biotin synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioB"
FT  /number= 1
FT  1141..1146
FT  /tag= f
FT  /standard_name= "bioF RBS"
FT  1154..2311
FT  /tag= g
FT  /EC_number= 2.3.1.47
FT  /product= "KAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioF"
FT  /number= 2
FT  /standard_name= "8-amino-7-oxononanoate synthase"
FT  2284..2288
FT  /tag= h
FT  /standard_name= "bioC RBS"
FT  2295..3050
FT  /tag= i
FT  /function= "involved in pimeoyl-CoA synthesis"
FT  /product= "protein"
FT  /gene= "bioC"
FT  /number= 3
FT  3030..3033
FT  /tag= j
FT  /standard_name= "bioD RBS"
FT  3043..3753
FT  /tag= k
FT  /EC_number= 6.3.3.3
FT  /product= "DTB synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioD15"
FT  /number= 4
FT  /standard_name= "dehydrobiotin synthase"
FT  3712..3750
FT  /tag= l
FT  /note= "bioD15 substitution"
FT  3742..3746
RBS
misc_RNA
RBS

```

```

FT  /tag= m
FT  /standard_name= "bioA RBS"
FT  3750..5039
FT  /tag= n
FT  /EC_number= 2.6.1.62
FT  /product= "DAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioA"
FT  /number= 5
FT  /standard_name= "S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase"
FT  5088..5093
FT  /tag= o
FT  /standard_name= "ORF1 RBS"
FT  5098..5574
FT  /tag= p
FT  /function= "unknown, involved in biotin synthesis"
FT  /product= "protein"
FT  /evidence= EXPERIMENTAL
FT  /gene= "ORF1"
FT  /number= 6
FT  5583..5644
FT  /tag= q
FT  /standard_name= "rho-independent transcriptional terminator"
FT  5583..5605
FT  /tag= r
FT  stem_loop
FT  5583..5605
FT  /tag= r

```

```

W09408023-A.
14-APR-1994.
01-OCT-1993; 93WO-EP02688.
02-OCT-1992; 92CH-0003124.
15-JUL-1993; 93CH-0002134.
(LONZA ) LONZA AG.
Birch O, Brass J, Fuhrmann M, Shaw N;
WPI; 1994-135587/16.
P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
Biotechnological biotin prodn. using enterobacterial biotin-gene
- providing vitamin H in high yield
Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
The sequence is derived from plasmid pB030A-15/9 contg. the
bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
of biotin, arranged in a transcription unit. Microorganisms
contg. these DNA fragments or plasmids may be used in the prodn.
of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
loss of appetite and tiredness.
Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

```

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtgtgtgattcgtcgtgttgaacg 25
    |||||||
Db 173 gtgtgtgattcgtcgttgaacg 197

```

```

RESULT 6
AAT49104
ID AAT49104 standard; DNA; 456 BP.
XX
AC AAT49104;

```

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 XX Patton DA;
 XX
 XX WPI: 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 XX Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX
 XX Example 2: Column 37-40; 34pp: English.
 PS
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SO Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtgtgatactgctgttgaagcg 25
 ||||||||||||||||||||||||
 Db 57 gtgtgtgatactgctgttgaagcg 81

RESULT 3
 AAN91329
 ID AAN91329 standard; DNA; 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag-a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 PS Table 3; page 33-4; 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMAB1, pMAB6c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SO Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtgtgatactgctgttgaagcg 25
 ||||||||||||||||||||||||
 Db 80 gtgtgtgatactgctgttgaagcg 104

RESULT 4
 AAN60496
 ID AAN60496 standard; DNA; 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag= a
 XX
 PN JP61149091-A.
 XX
 ED 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 XX
 SO Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtgtgatactgctgttgaagcg 25
 ||||||||||||||||||||||||
 Db 98 gtgtgtgatactgctgttgaagcg 122

RESULT 5
 AAO62386
 ID AAO62386 standard; DNA; 5872 BP.
 XX


```

C 85 16.6 66.4 3603 17 AA142217 Human TATA-binding
C 86 16.6 66.4 3603 18 AA179595 TATA-binding prote
C 87 16.6 66.4 5321 21 AAA29207 Human DEX1 coding
C 88 16.6 66.4 5322 19 AA169631 Dead Box X (DBX) g
C 89 16.6 66.4 5338 20 AA131242 Enterococcus faeca
C 90 16.4 65.6 336 22 AA165487 Novel human polynu
C 91 16.4 65.6 1061 10 AA191022 Rotavirus gene 9 e
C 92 16.4 65.6 3731 22 AA104370 Tomato RNA-directe
C 93 16.2 64.8 74 21 AAC29046 Human secreted pro
C 94 16.2 64.8 149 22 AA125552 Probe #15485 for g
C 95 16.2 64.8 149 22 AA152082 Probe #20768 used
C 96 16.2 64.8 311 22 AA170876 Human cervical can
C 97 16.2 64.8 311 22 AA172433 Human cervical can
C 98 16.2 64.8 338 22 AA169541 Human cervical can
C 99 16.2 64.8 464 22 AA170077 Human cervical can
C 100 16.2 64.8 467 22 AA116371 Probe #6304 for ge

```

ALIGNMENTS

```

RESULT 1
AA162941
ID AA162941 standard; DNA; 839 BP.
XX
AC AA162941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SRO ID NO 269.
XX
XX Human; noctropic; neuroprotective; cytoskeletal; dermatological; virocidic;
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnereary;
XX antiparkinsonian; antischistosomal; antianemic; antidiabetic; cancer;
XX antiallergic; hepatotropic; cerebroprotective; anti-inflammatory;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX ds.
XX
XX Homo sapiens.
XX
XX WO200155449-A1.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01346.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 19-MAY-2000; 2000US-020515.
XX 07-JUL-2000; 2000US-0216880.
XX 14-JUL-2000; 2000US-0218290.
XX 14-AUG-2000; 2000US-0225447.
XX 01-SEP-2000; 2000US-0229343.
XX 06-SEP-2000; 2000US-0230437.
XX 08-SEP-2000; 2000US-0231283.
XX 25-SEP-2000; 2000US-0234997.
XX 29-SEP-2000; 2000US-0236367.
XX 13-OCT-2000; 2000US-0239937.
XX 08-NOV-2000; 2000US-0246476.
XX 08-NOV-2000; 2000US-0246477.
XX 08-NOV-2000; 2000US-0246477.
XX 08-NOV-2000; 2000US-0246525.
XX 08-NOV-2000; 2000US-0246526.
XX 08-NOV-2000; 2000US-0246528.
XX 17-NOV-2000; 2000US-0249210.
XX 17-NOV-2000; 2000US-0249211.
XX 17-NOV-2000; 2000US-0249214.
XX 17-NOV-2000; 2000US-0249265.
XX 01-DEC-2000; 2000US-0250160.
XX 01-DEC-2000; 2000US-0250391.
XX 05-DEC-2000; 2000US-0251030.
XX 05-DEC-2000; 2000US-0251988.

```

```

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM.
XX
XX WPI; 2001-476225/51.
XX
XX Novel plasma membrane associated proteins useful for diagnosing,
XX treating, preventing and/or prognosing disorders related to the
XX proteins, including cancer, immune response and neuronal disorders
XX
XX
XX Example 2: SEQ ID NO 269; 532pp + Sequence listing; English.
XX
XX The invention relates to novel genes (AA162752-AA162961) and proteins
XX (AA162347-AA162415) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;
XX
XX
XX Query Match 100.0%; Score 25; DB 22; Length 839;
XX Best Local Similarity 100.0%; Pred. NO. 0.14;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 gtgcgcgacgcgtgttgaagcg 25
XX |||||||||||||||||||
XX Db 89 gtgcgcgacgcgtgttgaagcg 113
XX
XX
XX RESULT 2
XX AA01303
XX ID AA01303 standard; DNA; 1041 BP.
XX
XX
XX AC AA01303;
XX
XX
XX 12-APR-1999 (first entry)
XX
XX
XX E. coli biotin synthetase (BioB) coding sequence.
XX
XX
XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
XX biotin synthase; biotin production; vitamin H; BioB; ss.
XX
XX Escherichia coli.
XX
XX
XX OS US5869719-A.
XX
XX
XX PN 09-FEB-1999.
XX
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 08-MAR-1995; 95US-0401068.
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:17:41 : Search time 366.42 seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196f-3
Perfect score: 25
Sequence: 1 gttgctgagctgcgtcttgaagcg 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

N_Geneseq.1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli biotin syn
3	25	100.0	1084	10	AA01329 E. coli Bio B gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesi
6	17.6	70.4	456	18	AA049104 Partial DNA clone
7	17.6	70.4	527	21	AA055976 Eucaulypus grandis
8	17.6	70.4	774	21	AA053871 Neisseria meningit
9	17.6	70.4	1236	22	AA064991 C glutamicum codin
10	17.6	70.4	1359	22	AA072295 Corynebacterium gl
11	17.6	70.4	3043	21	AA078167 Human cancer assoc

12	17.6	70.4	4573	20	AA033946
13	17.6	70.4	4610	22	AA059633
14	17.6	70.4	52253	21	AA081478
15	17.6	70.4	349980	21	AA021544
16	17.6	70.4	349980	22	AA064966
17	17.6	70.4	1437668	21	AA081480
18	17.4	69.6	1994	21	AA021149
19	17.4	69.6	1994	21	AA081426
20	17.4	69.6	1994	21	AA035027
21	17.4	69.6	2009	20	AA020513
22	17.4	69.6	2034	20	AA020751
23	17.4	69.6	8631	21	AA021150
24	17.4	69.6	8631	21	AA035028
25	17.2	68.8	434	21	AA030560
26	17.2	68.8	687	21	AA013715
27	17.2	68.8	1068	21	AA029542
28	17.2	68.8	1250	22	AA060712
29	17.2	68.8	1317	22	AA058926
30	17.2	68.8	10564	16	AA010854
31	17.2	68.8	14822	20	AA020543
32	17.2	68.8	1230025	20	AA031990
33	17	68.0	157	21	AA012134
34	17	68.0	272	21	AA087515
35	17	68.0	492	21	AA044161
36	17	68.0	729	20	AA087783
37	17	68.0	751	21	AA058914
38	17	68.0	826	22	AA0806423
39	17	68.0	925	20	AA088879
40	17	68.0	1008	18	AA070131
41	17	68.0	1200	21	AA048910
42	17	68.0	1237	21	AA034154
43	17	68.0	1514	20	AA039643
44	17	68.0	1573	21	AA048922
45	17	68.0	1574	21	AA037600
46	17	68.0	2709	22	AA016213
47	17	68.0	2750	22	AA088117
48	17	68.0	3029	22	AA0160452
49	17	68.0	3435	17	AA035869
50	17	68.0	3845	22	AA058666
51	17	68.0	4072	21	AA025054
52	17	68.0	7117	22	AA003131
53	16.8	67.2	309	20	AA086759
54	16.8	67.2	669	22	AA050264
55	16.8	67.2	1328	19	AA058754
56	16.8	67.2	12284	11	AA006001
57	16.8	67.2	12284	15	AA021308
58	16.8	67.2	12284	17	AA025591
59	16.8	67.2	12284	22	AA030678
60	16.8	67.2	63563	22	AA028546
61	16.8	67.2	611590	21	AA022303
62	16.6	66.4	307	21	AA011797
63	16.6	66.4	417	20	AA051612
64	16.6	66.4	417	20	AA000466
65	16.6	66.4	437	20	AA040715
66	16.6	66.4	588	22	AA013655
67	16.6	66.4	749	19	AA053606
68	16.6	66.4	1068	21	AA030675
69	16.6	66.4	1068	21	AA030743
70	16.6	66.4	1161	19	AA018352
71	16.6	66.4	1162	15	AA066170
72	16.6	66.4	1233	21	AA029206
73	16.6	66.4	1272	22	AA052048
74	16.6	66.4	1694	21	AA030329
75	16.6	66.4	2093	17	AA053720
76	16.6	66.4	2093	17	AA011864
77	16.6	66.4	2116	22	AA018629
78	16.6	66.4	2214	15	AA070731
79	16.6	66.4	2254	15	AA066167
80	16.6	66.4	2254	19	AA018343
81	16.6	66.4	2254	19	AA018343
82	16.6	66.4	2561	21	AA075822
83	16.6	66.4	3119	21	AA091717
84	16.6	66.4	3408	19	AA069632

Human HCMV inducib
Human cell cycle a
N. meningitidis pa
Neisseria meningit
C glutamicum codin
N. meningitidis B
Human low adenosin
Human I-kappa-B ki
Human adenosine re
Human RIP-associat
Human low adenosin
Human adenosine re
Human secreted pro
Aspergillus oryzae
Cyclophilin-type p
Human polynucleoti
Human polynucleoti
Human fibrinogen g
Polynucleotide seq
Nucleotide sequenc
Human secreted pro
Rat hepatocyte car
Arabidopsis thalia
EST clone BS306.
DNA encoding ketos
Human cDNA clone (
EST clone H2103.
Max-interacting pr
Arabidopsis thalia
Arabidopsis thalia
Renal cancer assoc
Arabidopsis thalia
Arabidopsis thalia
Human cDNA sequenc
Human FLEXN-48 nu
Human polynucleoti
Human DNA polymera
Human polynucleoti
Arabidopsis thalia
COL3-CA8 vector
EST clone AX318.
Human secreted pro
Human secreted pro
Sequence encoding
Hog cholera virus.
Hog cholera virus
Genomic fragment #
Arabidopsis thalia
Aspergillus niger
Human secreted pro
Human secreted pro
Human secreted pro
Probe #8341 used t
Group B Streptococ
Human G protein-co
DNA encoding human
Human V28 seven tr
Seven transmembran
Human topoisomeras
Mycobacterium tube
Flax 2S storage pr
Arabidopsis thalia
Hepatitis A virus
Human cDNA sequenc
tRNA-binding prote
Seven transmembran
Human V28 seven tr
Human V28 receptor
Human ORF3 ORF137
Human 7TM receptor
Dead Box X (DBX) g

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-392P13"
/clone_lib="RP11-392P13"
BASE COUNT 83194 a 47434 c 46923 g 82192 t 2865 others
ORIGIN

Query Match 73.6% Score 18.4; DB 2; Length 262608;
Best Local Similarity 95.0%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ttgctgagctcgtgttga 21
|||||
Db 217049 TTGCTGATCTGTGTGTA 217030

RESULT 45
AE002676 2018 bp DNA INV 06-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 14200001385608, complete
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2018)

REFERENCE
AUTHORS

Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.C., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,T.H., Blazey,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,M., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guen,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpis,G.H., Ke,Z.,
Kensler,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matley,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Modarres,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhou,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006

REFERENCE 2 (bases 1 to 2018)
AUTHORS Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gi:7289371.
FEATURES
Location/Qualifiers
source
1..2018
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 473 a 442 c 424 g 679 t
ORIGIN

Query Match 72.8% Score 18.2; DB 3; Length 2018;
Best Local Similarity 87.0%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgctgagctcgtgttgaac 24
|||||
Db 1294 TTGCTGATCTGCTGTGTGCGC 1316

Search completed: December 26, 2001, 10:56:03
Job time: 1925 sec

* 128571 136687: contig of 8117 bp in length
* 136688 136787: gap of unknown length
* 136788 144774: contig of 7987 bp in length
* 144775 144874: gap of unknown length
* 144875 152942: contig of 8068 bp in length
* 152943 153042: gap of unknown length
* 153043 164737: contig of 11695 bp in length
* 164738 164837: gap of unknown length
* 164838 174879: contig of 10042 bp in length
* 174880 174979: gap of unknown length
* 174980 184262: contig of 9283 bp in length
* 184263 184362: gap of unknown length
* 184363 195368: contig of 11006 bp in length
* 195369 195468: gap of unknown length
* 195469 205038: contig of 9570 bp in length
* 205039 205138: gap of unknown length
* 205139 217691: contig of 12553 bp in length
* 217692 229573: gap of unknown length
* 229574 229674: gap of unknown length
* 229675 242015: contig of 12342 bp in length.
Location/Qualifiers
1. .242015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-195A20"
/clone_1lb="RPC1 human BAC library 11"

BASE COUNT 75287 a 43117 c 43007 g 75723 t 4881 others
ORIGIN

Query Match 73.6% Score 18.4; DB 2; Length 242015;
Best Local Similarity 95.0% Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgcgtgacgcgtgttga 21
|||||
Db 80568 TTGCTGATCTGTTTGA 80569

RESULT 44
AC092362/c
LOCUS
DEFINITION HOMO sapiens chromosome 5 clone RP11-392P13, WORKING DRAFT
ACCESSION AC092362
VERSION AC092362.1 GI:14589551
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 262608)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 262608)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov

Project Information
Center Project Name: 557266
Center clone name: RPC1-11_392P13

Summary Statistics
Consensus quality: 238118 bases at least Q40

Consensus quality: 252568 bases at least Q30
Consensus quality: 254471 bases at least Q20
Estimated insert size: 167930; agarose-fp estimation
Estimated insert size: 259808; sum-of-contigs estimation
Quality coverage: 14.89 in Q20 bases; agarose-fp estimation
Quality coverage: 9.62 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a "working draft" sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1473: contig of 1473 bp in length
1474 1573: gap of unknown length
1574 2720: contig of 1147 bp in length
2721 2820: gap of unknown length
2821 4142: contig of 1322 bp in length
4143 4242: gap of unknown length
4243 5814: contig of 1572 bp in length
5815 5914: gap of unknown length
5915 6987: contig of 1073 bp in length
6988 7087: gap of unknown length
7088 8448: gap of 1361 bp in length
8449 8548: gap of unknown length
8549 9583: contig of 1035 bp in length
9584 9683: gap of unknown length
9684 11798: contig of 2115 bp in length
11799 11898: gap of unknown length
11899 14139: contig of 2241 bp in length
14140 14239: gap of unknown length
14240 16720: contig of 2481 bp in length
16721 16820: gap of unknown length
16821 23149: contig of 6339 bp in length
23150 23249: gap of unknown length
23250 29355: contig of 6106 bp in length
29356 29455: gap of unknown length
29456 35681: contig of 6226 bp in length
35682 35781: gap of unknown length
35782 42019: contig of 6238 bp in length
42020 42119: gap of unknown length
42120 49804: contig of 7665 bp in length
49805 49904: gap of unknown length
49905 58842: contig of 8938 bp in length
58843 58942: gap of unknown length
58943 68221: contig of 9279 bp in length
68222 68321: gap of unknown length
68322 78413: contig of 10092 bp in length
78414 78513: gap of unknown length
78514 90243: contig of 11730 bp in length
90244 90343: gap of unknown length
90344 105334: contig of 14991 bp in length
105335 105434: gap of unknown length
105435 115033: contig of 9599 bp in length
115034 115133: gap of unknown length
115134 122391: contig of 7258 bp in length
122392 122491: gap of unknown length
122492 129865: contig of 7374 bp in length
129866 129965: gap of unknown length
129966 144625: contig of 14660 bp in length
144626 144725: gap of unknown length
144726 165714: contig of 20989 bp in length
165715 165814: gap of unknown length
165815 186267: contig of 20453 bp in length
186268 186367: gap of unknown length
186368 199645: contig of 13278 bp in length
199646 199745: gap of unknown length
199746 219844: contig of 20099 bp in length
219845 219944: gap of unknown length
219945 262608: contig of 42664 bp in length.
Location/Qualifiers
1. .262608

FEATURES
source

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0347008

FEATURES
source
1. 193980
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-34708"
BASE COUNT 58718 a 36662 c 37932 g 60668 t
ORIGIN
Query Match 73.6% Score 18.4; DB 9; Length 193980;
Best Local Similarity 95.0%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 3 tgcgcgactctctgttga 22
|||||
Db 96921 TGCTGATCTCTTTTGA 96940

RESULT 43
AC092333/c
LOCUS
DEFINITION HMO sapiens chromosome 5 clone RP11-195A20, WORKING DRAFT
SEQUENCE, 48 unordered pieces.
ACCESSION AC092333 GI:14589522
VERSION AC092333.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 242015)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 242015)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 481265
Center clone name: RPCI-11_195A20

Summary Statistics
Consensus quality: 193910 bases at least Q40
Consensus quality: 224362 bases at least Q30
Consensus quality: 228902 bases at least Q20
Estimated insert size: 158340; agarose-fp estimation
Estimated insert size: 237315; sum-of-contigs estimation
Quality coverage: 6.04 in Q20 bases; agarose-fp estimation
Quality coverage: 4.03 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1074: contig of 1074 bp in length

1075 1174: gap of unknown length
1175 2345: contig of 1171 bp in length
2346 2445: gap of unknown length
2446 3480: contig of 1035 bp in length
3481 3580: gap of unknown length
3581 4867: contig of 1287 bp in length
4868 4967: gap of unknown length
4968 6502: contig of 1535 bp in length
6503 7839: gap of unknown length
7840 7939: gap of unknown length
7940 9190: contig of 1251 bp in length
9191 9291: gap of unknown length
9291 10998: contig of 1708 bp in length
10999 11098: gap of unknown length
11099 13106: contig of 2008 bp in length
13107 13206: gap of unknown length
13207 16139: contig of 2933 bp in length
16140 16239: gap of unknown length
16240 17477: contig of 1238 bp in length
17478 17577: gap of unknown length
17578 18893: contig of 1416 bp in length
18894 19093: gap of unknown length
19094 20665: contig of 1572 bp in length
20666 20765: gap of unknown length
20766 23023: contig of 2258 bp in length
23024 23123: gap of unknown length
23124 24951: contig of 1828 bp in length
24952 25051: gap of unknown length
25052 27411: contig of 2360 bp in length
27412 27511: gap of unknown length
27512 30587: contig of 3076 bp in length
30588 30687: gap of unknown length
30688 32731: contig of 2044 bp in length
32732 32831: gap of unknown length
32832 34701: contig of 1870 bp in length
34702 34801: gap of unknown length
34802 38296: contig of 3495 bp in length
38297 38396: gap of unknown length
38397 41546: contig of 3150 bp in length
41547 41646: gap of unknown length
41647 44317: contig of 2671 bp in length
44318 44417: gap of unknown length
44419 46461: contig of 2044 bp in length
46462 46561: gap of unknown length
46562 49475: contig of 2914 bp in length
49476 49575: gap of unknown length
49576 51975: contig of 2400 bp in length
51976 52075: gap of unknown length
52076 54564: contig of 2489 bp in length
54565 54664: gap of unknown length
54665 57355: contig of 2691 bp in length
57356 57455: gap of unknown length
57456 65678: contig of 8223 bp in length
65679 65778: gap of unknown length
65779 72658: contig of 6880 bp in length
72659 72758: gap of unknown length
72759 77127: contig of 4369 bp in length
77128 77227: gap of unknown length
77228 83304: contig of 6077 bp in length
83305 83404: gap of unknown length
83406 88866: contig of 5462 bp in length
88867 88966: gap of unknown length
88967 97138: contig of 8172 bp in length
97139 97238: gap of unknown length
97239 102438: contig of 5200 bp in length
102439 102538: gap of unknown length
102539 112046: contig of 9508 bp in length
112047 112146: gap of unknown length
112147 120360: contig of 8214 bp in length
120361 120460: gap of unknown length
120461 128470: contig of 8010 bp in length
128471 128570: gap of unknown length

```

/exon 90256..90301
/gene="NFKB1"
/number=2
join(90253..90301,94580..94658,98592..98629,
102602..102700,131731..131879,141620..141783,
143625..143783,145279..145383,147610..147701,
149426..149564,158169..158312,159636..159725,
160882..161076,162264..162405,165639..165753,
171240..171441,171894..172063,172495..172633,
175319..175443,176788..176854,177178..177350,
178169..178325,181178..181338)
/gene="NFKB1"
/feature="p105; nuclear factor kappa-B DNA binding subunit"
/codon_start=1
/product="nuclear factor of kappa light polypeptide gene
enhancer in B-cells 1"
/protein_id="AF55232.1"
/db_xref="GI:7012906"
/translation="MAEDDYIGRPDMFHLDPSTLHTTFNPEVPOFQALPTDGPYI.
QILEOPKORGFRRFYVCEGSLPGASSEKRSYPQVKNVGPAPKAVIYQVLTN
GKNTLHLSLVGKHCEDGICVTAGKDMVGFANILHTTKKVFETLEARMTEA
CIRGYNGLIHLVDIAYLQAGGGDROLGDRKELRQALQOTKEMDLSYRLMFTA
PLDSTGSPTRRLRPVYSDAIVSKAPNANLKIYVMDRAGSVTGGSEYILICDVO
KDLQTFYEEENGWYEGGDFSPFDVHRQAIYFKTIPTKIDINTKPAFYVOLR
RKSDLETSEPRKPLYPETKDEVOKRQKLMPSNDSFGSGAGAGGGGSGG
GGGGTGTGPGVSPHYGPPTVGGITFHPGCTSNAGMKHMDTESKKDPECCDSD
DKNTVNLFGKVIETTEDODPESEATVGNCEVTLTYATGTEESAGVQDNLFKAKQL
AKRHANALPDYAVTGDKMLAVORHLTAQDENGDSVLHAIILHLSOLVRDLLEVY
SGILSDLIIMRNDLYOTPLHLAVITKQPDVENDLIRAGADSLDRIGNSVHLIAK
EGHDKVSLILKHKKALLDHPNGGLNLIHLMMSNSLPCILLVYAGADVNADO
KSGRTALHVAEHNDNISLACGLLEGAHVDSTTGGTPLHLIAAGSRTRIALKLA
AGADPLENEPEPLVDLDDSWENAGDEGVVPTTPIDMAISWGFVILNCKVPEPFT
SDDLAAGDKMQLAEDVKQLQYLKLEIPDPDKMWAALQAKLGIILNMFRLSPAPSK
TLMNDYEVSGTVAELVLRQMGYTEALVIAASSPVKTSQASHLPLSPASTRQO
IDELRDSYCDGCVETSFRLKSTFSTSLTSGASLTLNKKMHPHYGGEGPLEGKI"
/exon 94580..94658
/gene="NFKB1"
/number=3
98592..98629
/gene="NFKB1"
/number=4
102602..102700
/gene="NFKB1"
/number=5
131731..131879
/gene="NFKB1"
/number=6
141620..141783
/gene="NFKB1"
/number=7
143625..143783
/gene="NFKB1"
/number=8
145279..145383
/gene="NFKB1"
/number=9
147610..147701
/gene="NFKB1"
/number=10
149426..149564
/gene="NFKB1"
/number=11
158169..158312
/gene="NFKB1"
/number=12
159636..159725
/gene="NFKB1"
/number=13
160882..161076
/gene="NFKB1"
/number=14
162264..162405

```

```

/exon 156639..165753
/gene="NFKB1"
/number=15
171240..171441
/gene="NFKB1"
/number=16
171894..172063
/gene="NFKB1"
/number=17
172393..172495
/gene="NFKB1"
/number=18
175319..175443
/gene="NFKB1"
/number=19
176788..176854
/gene="NFKB1"
/number=20
176788..176854
/gene="NFKB1"
/number=21
177178..177350
/gene="NFKB1"
/number=22
178169..178325
/gene="NFKB1"
/number=23
181178..181659
/gene="NFKB1"
/number=24
189901..190000
/feature="note="sequence overlaps with 5' end of sequence in
Genbank Accession Number AF224669"

```

Query Match 73.6%; Score 18.4; DB 9; Length 190000;
Best Local Similarity 95.0%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcgtgacgtcgtgttga 22
|||||
Db 127950 TGCTGATCTCGCTTTTGA 127931

RESULT 42
AC021120
LOCUS AC021120 193980 bp DNA PRI 14-JUL-2001
DEFINITION Homo sapiens clone RP11-34708, complete sequence.
AC021120
ACCESSION AC021120.6 GI:14718391
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 193980)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193980)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 193980)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 14, 2001 this sequence version replaced gi:14573743.

```

* 33126 33225: gap of unknown length
* 33226 37707: contig of 4482 bp in length
* 37708 37807: gap of unknown length
* 37808 43473: contig of 5666 bp in length
* 43474 43573: gap of unknown length
* 43574 46111: contig of 6038 bp in length
* 46112 49712: gap of unknown length
* 49712 58177: contig of 8465 bp in length
* 58177 58277: gap of unknown length
* 58277 65583: contig of 7306 bp in length
* 65583 73088: gap of unknown length
* 73088 73189: contig of 7406 bp in length
* 73189 82463: gap of unknown length
* 82463 91393: contig of 9274 bp in length
* 91393 91493: gap of unknown length
* 91493 103163: contig of 8830 bp in length
* 103164 103263: gap of unknown length
* 103264 115815: contig of 11671 bp in length
* 115816 115915: gap of unknown length
* 115916 131166: contig of 12552 bp in length
* 131167 131267: gap of unknown length
* 131267 147652: gap of unknown length
* 147652 147752: contig of 16385 bp in length
* 147752 168091: gap of unknown length
* 168091 20340: contig of 20340 bp in length.
Location/Qualifiers
1. 168091
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-27P9"
1. 1062
/note="assembly_name:Contig15"
misc_feature
1163. 2457
/note="assembly_name:Contig6"
misc_feature
2558. 4705
/note="assembly_name:Contig7"
misc_feature
4806. 6688
/note="assembly_name:Contig8"
misc_feature
6789. 8449
/note="assembly_name:Contig9"
misc_feature
clone_end:77
vector_side:left
8550. 9683
/note="assembly_name:Contig10"
misc_feature
9784. 11912
/note="assembly_name:Contig11"
misc_feature
12013. 15816
/note="assembly_name:Contig12"
misc_feature
15917. 18597
/note="assembly_name:Contig13"
misc_feature
18698. 22258
/note="assembly_name:Contig14"
misc_feature
22359. 24779
/note="assembly_name:Contig15"
misc_feature
24880. 27994
/note="assembly_name:Contig16"
misc_feature
28095. 33125
/note="assembly_name:Contig17"
misc_feature
33226. 37707
/note="assembly_name:Contig18"
misc_feature
37808. 43473
/note="assembly_name:Contig19"
misc_feature
43574. 49611
/note="assembly_name:Contig20"
misc_feature
49712. 58176
/note="assembly_name:Contig21"
misc_feature
58277. 65582
/note="assembly_name:Contig22"
misc_feature
65683. 73088
/note="assembly_name:Contig23"
misc_feature
73189. 82462

```

```

/note="assembly_name:Contig24"
82563. 91392
/note="assembly_name:Contig25"
91493. 103163
/note="assembly_name:Contig26"
103264. 115815
/note="assembly_name:Contig27"
115916. 131166
/note="assembly_name:Contig28"
131267. 147651
/note="assembly_name:Contig29"
clone_end:SP6
vector_side:right"
147752. 168091
/note="assembly_name:Contig30"
misc_feature
51267 a 31894 c 31107 g 51315 t 2508 others
BASE COUNT
ORIGIN
Query Match 73.6% Score 18.4; DB 2; Length 168091;
Best Local Similarity 95.0%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ttgctgacatcgtctgttga 21
|||||
Db 144862 ttgctgacatcgtctgttga 144843

```

```

RESULT 41
AF213884S1/c AF213884S1 190000 bp DNA PRI 21-FEB-2000
LOCUS Homo sapiens nuclear factor of kappa light polypeptide gene
DEFINITION enhancer in B-cells 1 (NFkB1) gene, complete cds.
ACCESSION AF213884
VERSION AF213884.1 GI:7012904
KEYWORDS 1 of 2
SEGMENT human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 190000)
AUTHORS Chang, H.-M. and Tsai, S.-F.
TITLE Genome Sequencing of the Chromosome 4q Region Implicated in Human
Hepatocellular Carcinoma Pathogenesis
JOURNAL Unpublished
2 (bases 1 to 190000)
AUTHORS Chang, H.-M. and Tsai, S.-F.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Institute of Genetics, National Yang-Ming
University, 155 Li-Hong St. Section 2, Peitou, Taipei, Taiwan
11221, Republic of China
Location/Qualifiers
1. 190000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q"
66142. 66532
/feature="NFkB1"

```

```

FEATURES
source
1. 190000
/number=1
join(66142..66532,90256..90301,94580..94658,98592..98629,
102602..102700,131731..131879,141620..141783,
143625..143783,145279..145383,147610..147701,
149426..149564,158169..158312,159636..159725,
160882..161076,162264..162405,165639..165733,
171240..171441,171894..172063,172393..172493,
175319..175443,176788..176854,177178..177350,
178169..178325,181178..181659)
/feature="NFkB1"
/product="nuclear factor of kappa light polypeptide gene
enhancer in B-cells 1"
66142..181659
gene

```

```

/gene="pyk20"
/citation="[1]
/function="nematode responsive gene"
BASE COUNT      1286 a      638 c      658 g      1455 t
ORIGIN

Query Match      74.4%: Score 18.6; DB 8; Length 4037;
Best Local Similarity 84.0%: Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 gttgctgagatcgtctgttgaagcg 25
        ||||| ||||| ||||| |||||
Db      533 GTTGCTGAATCTGCTGTTGAGCG 509

RESULT 39
AC034199      122877 bp      DNA      PRI      29-SEP-2000
LOCUS
DEFINITION    Homo sapiens chromosome 5 clone CTB-114C7, complete sequence.
ACCESSION     AC034199
VERSION       AC034199.5 GI:10337637
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE     DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS       Unpublished
TITLE         Direct Submission
JOURNAL       2 (bases 1 to 122877)
REFERENCE     DOE Joint Genome Institute.
AUTHORS       Direct Submission
TITLE         Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
JOURNAL       Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
               3 (bases 1 to 122877)
REFERENCE     DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS       Direct Submission
TITLE         Submitted (29-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL       Drive, Walnut Creek, CA 94598, USA
               On Sep 29, 2000 this sequence version replaced gi:9256716.
               Draft Sequence Produced by DOE Joint Genome Institute
               www.jgi.doe.gov
               Finishing Completed at Stanford Human Genome Center
               www.sngc.stanford.edu
               Quality: Phrap Quality >=40 99.5% of Sequence;
               Estimated Total Number of Errors is 0.4.
               STS Content:
               SHGC-104344 G58254.
FEATURES
Source        Location/Qualifiers
               1..122877
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="5"
               /clone="CTB-114C7"
BASE COUNT    33261 a 27391 c 27395 g 34830 t
ORIGIN

Query Match      74.4%: Score 18.6; DB 9; Length 122877;
Best Local Similarity 84.0%: Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 gttgctgagatcgtctgttgaagcg 25
        ||||| ||||| ||||| |||||
Db      57185 GTTGCTGAATCTGCTTTACAAGG 57209

RESULT 40
AC016895/c      168091 bp      DNA      HTG      07-JUL-2000
LOCUS
DEFINITION      Homo sapiens chromosome 5 clone RP11-27P9, WORKING DRAFT SEQUENCE,

```

```

ACCESSION      AC016895
VERSION         AC016895.2 GI:7630807
KEYWORDS        HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE          human.
ORGANISM        Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
               1 (bases 1 to 168091)
REFERENCE       Waterston, R.H.
AUTHORS         The sequence of Homo sapiens clone
TITLE           Unpublished
JOURNAL         2 (bases 1 to 168091)
REFERENCE       Waterston, R.H.
AUTHORS         Direct Submission
TITLE           Submitted (08-DEC-1999) Genome Sequencing Center, Washington
JOURNAL         University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63109, USA
               On Apr 21, 2000 this sequence version replaced gi:6539412.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0027P09
----- Summary Statistics -----
Sequencing vector: M13; 74%
Sequencing vector: plasmid; 26%
Chemistry: Dye-primer ET; 74% of reads
Chemistry: Dye-terminator Big Dye; 26% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 153848 bases at least Q40
Consensus quality: 158311 bases at least Q30
Consensus quality: 160576 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 165591; sum-of-contigs
Quality coverage: 3.47 in Q20 bases; agarose-fp
Quality coverage: 3.49 in Q20 bases; sum-of-contigs
-----

```

```

NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1063      1062: contig of 1062 bp in length
* 1163      1162: gap of unknown length
* 1258      2457: contig of 1295 bp in length
* 2458      2557: gap of unknown length
* 2558      4705: contig of 2148 bp in length
* 4706      4805: gap of unknown length
* 4806      6688: contig of 1883 bp in length
* 6689      6788: gap of unknown length
* 6789      8449: contig of 1661 bp in length
* 8450      8549: gap of unknown length
* 8550      9683: contig of 1134 bp in length
* 9684      9783: gap of unknown length
* 9784      11912: contig of 2129 bp in length
* 11913      12012: gap of unknown length
* 12013      15816: contig of 3804 bp in length
* 15817      15916: gap of unknown length
* 15917      18597: contig of 2681 bp in length
* 18598      18697: gap of unknown length
* 18698      22258: contig of 3561 bp in length
* 22259      22358: gap of unknown length
* 22359      24779: contig of 2421 bp in length
* 24780      24879: gap of unknown length
* 24880      27994: contig of 3115 bp in length
* 27995      28094: gap of unknown length
* 28095      33125: contig of 5031 bp in length

```


TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165490)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 461289
Center clone name: RPCI-11_143A12

Summary Statistics
Consensus quality: 162787 bases at least Q40
Consensus quality: 164583 bases at least Q30
Consensus quality: 164855 bases at least Q20
Estimated insert size: 172290; agarose-1p estimation
Estimated insert size: 165390; sum-of-contigs estimation
Quality coverage: 7.7 in Q20 bases; agarose-1p estimation
Quality coverage: 8.02 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 38050: contig of 38050 bp in length
* 38051 38150: gap of unknown length
* 38151 165490: contig of 127340 bp in length.

FEATURES
source
1.165490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-143A12"
/clone.lib="RPCI human BAC library 11"
BASE COUNT 50499 a 31590 c 31648 g 51653 t 100 others
ORIGIN
Query Match 75.2%; Score 18.8; DB 2; Length 165490;
Best Local Similarity 90.9%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ttgcgtgacgtcgtttgaag 23
||||| ||||| ||||| |||||
Db 9622 ttgcgtgacgtcgtttgaag 9601

TITLE Shevchenko, Y., Snyder, B., Stantiriop, S., Thomas, J.W., Thomas, P.J.,
JOURNAL Tlionson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
REFERENCE Wetberly, K.D., Zhang, L.-H. and Green, E.D.
AUTHORS NISC Comparative Sequencing Initiative
JOURNAL Unpublished
TITLE 2 (bases 1 to 175977)
Submitted (24-MAY-2001) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
-----Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@nhgri.nih.gov

Project Information
Center project name: cea
Center clone name: 097119

Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174162 bases at least Q40
Consensus quality: 174943 bases at least Q30
Consensus quality: 175314 bases at least Q20
Insert size: 166000; agarose-1p
Insert size: 175677; sum-of-contigs
Quality coverage: 14.26x in Q20 bases; agarose-1p
Quality coverage: 13.48x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 8836: contig of 8836 bp in length
* 8837 8936: gap of unknown length
* 8937 23469: contig of 14533 bp in length
* 23470 23570: gap of unknown length
* 23570 107312: contig of 83743 bp in length
* 107313 107412: gap of unknown length
* 107413 175977: contig of 68565 bp in length.

FEATURES
source
1.175977
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-97L19"
/clone.lib="RP43"
1.8836
/note="assembly-fragment
clone-end:Sp6
vector-side:right"
8937 .23469
/note="assembly-fragment"
23570 .107312
/note="assembly-fragment"
107413 .175977
/note="assembly-fragment
clone-end:17
vector-side:left"
BASE COUNT 55853 a 35161 c 33286 g 51358 t 319 others
ORIGIN
Query Match 75.2%; Score 18.8; DB 2; Length 175977;
Best Local Similarity 90.9%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gtgcgtgacgtcgtttgaag 22
||||| ||||| ||||| |||||

```

/rpl_family="LINE/L1"
/rpl_type=DISPERSED
repeat_region 17451..19052
/note="LIMB8"
/rpl_family="LINE/L1"
/rpl_type=DISPERSED
repeat_region 19480..20289
/note="LIMB8"
/rpl_family="LINE/L1"
/rpl_type=DISPERSED
repeat_region complement(20290..20826)
/note="MER1A"
/rpl_family="DNA/MER1_type"
/rpl_type=DISPERSED
repeat_region 20827..21658
/note="LIMB8"
/rpl_family="LINE/L1"

Query Match 76.8%; Score 19.2; DB 9; Length 340000;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gtgtcgtgactcgtgttgaagc 24
||||| ||||||| |||||||
Db 150650 GATGCTGTTCTGCTTTGAAGC 150673

RESULT 33
AF158246 550 bp DNA ROD 12-JUN-2001
LOCUS Cricetulus griseus glucose phosphate isomerase (GPI) gene, partial
DEFINITION Intron sequence.
ACCESSION AF158246
VERSION AF158246.1 GI:5690369
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 550)
AUTHORS Williams,R.R.E., Hassan-Walker,A.F., Lavender,F.L., Morgan,M.,
Paik,P. and Ragoussis,J.
TITLE The minisatellite of the GPI/AMF/NLK/MF gene: interspecies
conservation and transcriptional activity
JOURNAL Gene 269 (1-2), 81-92 (2001)
MEDLINE 21272519
PUBMED 11376940
REFERENCE 2 (bases 1 to 550)
AUTHORS Williams,R.R.E., Hassan-Walker,A., Lavender,L., Morgan,M.M.J.,
Paik,P. and Ragoussis,J.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Medical and Molecular Genetics, King's
College London, Guy's Tower, Guy's Hospital, London SE1 9RT, UK
FEATURES
source
1..550
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
<1..>550
/gene="GPI"
<1..>550
/gene="GPI"
/note="glucose phosphate isomerase"
BASE COUNT 110 a 126 c 176 g 138 t
ORIGIN
Query Match 75.2%; Score 18.8; DB 10; Length 550;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtgtcgtgactcgtgttgaag 22
||||| ||||||| |||||||

```

```

Db 385 GTTCTGATCTGCTGTGAAG 406

RESULT 34
AC021088/c 106129 bp DNA PRI 21-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RPI1-143A12, WORKING DRAFT
DEFINITION AC021088
ACCESSION AC021088
VERSION AC021088.5 GI:14993686
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277275.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
source
1..106129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTD-214A45"
BASE COUNT 30092 a 20715 c 21187 g 34135 t
ORIGIN
Query Match 75.2%; Score 18.8; DB 9; Length 106129;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 ttgtcgtgactcgtgttgaag 23
||||| ||||||| |||||||
Db 48876 TTGCTGATCTGCTGTGAAG 48855

RESULT 35
AC091895/c 165490 bp DNA HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RPI1-143A12, WORKING DRAFT
DEFINITION AC091895
ACCESSION AC091895
VERSION AC091895.1 GI:14333831
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165490)
AUTHORS DOE Joint Genome Institute.

```

```

* Deutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizu@omb-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/.
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 48537
/note="Accession No. AF043945"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P39C17, 5' partial"
/clone_1lb="RPC11,3-5 PAC library"
1803. 2037
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(2442. 2739)
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
2906. 3301
/note="L1M3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
3565. 3599
/note="(TC)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3599. 3618
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3712. 3741
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(4324. 4371)
/note="MLT2CB"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
4415. 4446
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(4481. 4917)
/note="MLT2CB"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(5635. 5747)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
5972. 6057

```

```

/note="L1ME"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(6421. 6627)
/note="MER3"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
8410. 8615
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(9235. 9341)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(9523. 9962)
/note="MLT1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(9995. 10124)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10552. 10739)
/note="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
10970. 11104
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
12600. 13278
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
13318. 13662
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
13663. 13683
/note="(TTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(13684. 13968)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
13969. 14145
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
14320. 14340
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
14387. 14446
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(15002. 15087)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(15088. 15392)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(15393. 16150)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
16150. 17433
/note="L1PA16"

```

misc_feature /note="assembly_fragment"
31648..32297
misc_feature /note="assembly_fragment"
32398..33055
misc_feature /note="assembly_fragment"
33156..34038
misc_feature /note="assembly_fragment"
34139..34977
misc_feature /note="assembly_fragment"
35078..35765
misc_feature /note="assembly_fragment"
35866..36524
misc_feature /note="assembly_fragment"
36625..37591
misc_feature /note="assembly_fragment"
37692..38474
misc_feature /note="assembly_fragment"
38575..39518
misc_feature /note="assembly_fragment"
39619..40643
misc_feature /note="assembly_fragment"
40744..41393
misc_feature /note="assembly_fragment"
41494..42640
misc_feature /note="assembly_fragment"
42741..44211
misc_feature /note="assembly_fragment"
44312..45322
misc_feature /note="assembly_fragment"
45423..47108
misc_feature /note="assembly_fragment"
47209..48879
misc_feature /note="assembly_fragment"
48980..49882
misc_feature /note="assembly_fragment"
49983..51524
misc_feature /note="assembly_fragment"
51625..52638

Query Match 76.8%: Score 19.2; DB 2: Length 221420;
Best Local Similarity 87.5%: Pred. NO. 2.8e+02;

Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gttgctgagctgctgttgaagc 24
||||| ||||| ||||| |||||

Db 103303 GRTGCTGCTGCTGTTTGTGAC 103326

RESULT 31

LOCUS HSMX1A 300050 bp DNA PRI 09-MAY-2001
DEFINITION Homo sapiens chromosome 21 from 5 PACs and 5 Cosmids map
2122.2.D1S349-MX1: segment 1/2, complete sequence.

ACCESSION AL442166 AJ011929
VERSION AL442166.1 GI:10303259
KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 300050)
Ramsier, J., Francis, F., Beck, A., Hennig, S., Klages, S., Borzym, K.,

Langer, I., Steffens, C., Hildmann, T., Dagand, E., Yaspo, M.,
Reinhardt, R. and Lehnach, H.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 300050)

AUTHORS MPMG.

JOURNAL Direct Submission
Submitted (01-OCT-1998) MPMG, Abt. Lehnach, Max Planck Institut
fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
On May 9, 2001 this sequence version replaced gi:6982072.

COMMENT Clones received from Resource Centre of the Human Genome Project at
the Max-Planck-Institut for Molecular Genetics.

FEATURES Bases 300001..300050 overlap with AL442167.
Location/Qualifiers
1..300050

source /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"

/map="21q22.2.D1S349-MX1"

/clone="PAC RPCT-1 247E2"

/clone="PAC RPCT-1 146B4"

/clone="PAC RPCT-1 141D16"

/clone="PAC RPCT-1 269A14"

/clone="PAC RPCT-1 265B9"

/clone="cosmid LINC116 16H18"

/clone="cosmid LINC116 14C10"

/clone="cosmid LINC116 25D2"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

/clone="cosmid LINC116 87D5"

TITLE
JOURNAL
COMMENT

Baldwin, J., Barra, N., Beckerly, R., Boguslavsky, L., Bonkhalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArillano, K., Dewar, K., Domino, M., Donegan, D., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heiford, A., Horton, L., Howland, J. C., Johnson, S., Jones, C., Kann, L., Kariotas, A., Klein, J., Lenczky, J., Lien, C., Locke, K., McKernan, K., McQuinn, N., McGowan, P., Maygur, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severly, P., Stange, Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Tirelli, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 7, 2001 this sequence version replaced gi:5966274.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/M/RepeatMasker.html>

```

repeat_region /rpl_family="AT_rich"
complement(25703..25955)
/rpl_family="L1M6"
repeat_region 25956..26264
/rpl_family="AlusX"
complement(26265..26438)
/rpl_family="L1M6"
repeat_region 26863..26952
/rpl_family="L1M4"
repeat_region 26953..27441
/rpl_family="L1M4"
repeat_region 27442..27648
/rpl_family="L1M4"
repeat_region 28112..28304
/rpl_family="L1"
complement(28341..28631)
/rpl_family="Alusg"
repeat_region complement(28864..28933)
/rpl_family="L1M3"
repeat_region complement(29473..31095)
/rpl_family="L1M4"
repeat_region 31096..31127
/rpl_family="TTMAAn"
repeat_region complement(31128..31138)
/rpl_family="L1M4"
repeat_region complement(31139..31428)
/rpl_family="AluY"
repeat_region complement(31429..36250)
/rpl_family="L1M4"
repeat_region complement(36293..37911)
/rpl_family="L1M4"
repeat_region 38687..38708
/rpl_family="L1M6C"
repeat_region complement(39445..39934,40052..42824,49512..49607,
53304..53430,56142..56192)
/gene="ZNF228"
/product="ZNF228 mRNA"
/gene="ZNF228"
/complement(39445..56192)
/feature="zinc finger protein 228"
39524..39726

```

```

Query Match 76.8%; Score 19.2; DB 9; Length 176928;
Best Local Similarity 87.5%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 gtgtgcatctgtctgttgaacg 24
||||| ||||| ||||| |||||
Db 160901 GTGCTGCTGCTGCTTTTGTAGC 160878

```

```

RESULT 29
AC008224 181636 bp DNA INV 17-FEB-2001
LOCUS AC008224.c Drosophila melanogaster, chromosome 3R, region 83D-83D, BAC clone
DEFINITION BACR29J02, complete sequence.
ACCESSION AC008224
VERSION AC008224.18 GI:12957655
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```

```

REFERENCE
AUTHORS
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amandides,P.G., Brandon,R.C.,
Rogers,T., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Farihera,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

```

```

TITLE
JOURNAL
AUTHORS
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,R.A., Nunoo,J.,
Paclet,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phonemavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venier,J.C.
Sequencing of Drosophila chromosome 3R, region 83D-83D
Unpublished
2 (bases 1 to 181636)

```

```

TITLE
JOURNAL
AUTHORS
Celisner,S.E., Agbayani,A., Arcaluna,T.T., Baxter,E., Blazet,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomtan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclet,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shif,E.,
Svirkas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 17, 2001 this sequence version replaced gi:7239409.
COMMENT
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
location/Qualifiers

```

FEATURES

```

source
1..181636
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83D-83D"
/clone="BACR29J02 (DB17)"
/clone.lib="RP11-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6)"
BASE COUNT 51104 a 38903 c 38541 g 53088 t
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 3; Length 181636;
Best Local Similarity 87.5%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 ttgtgcatctgtctgttgaacg 25
||||| ||||| ||||| |||||
Db 148551 TTGCGGATGCTCTGTTTGAAGC 148528

```

```

RESULT 30
AC011822 221420 bp DNA HTG 07-MAY-2001
LOCUS AC011822 Homo sapiens chromosome 11 clone RP11-2022 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 37 unordered pieces.
ACCESSION AC011822
VERSION AC011822.6 GI:13959229
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 221420)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

```

TITLE Smith,L.M.
Sequence analysis of a 1mb region in 19q13.2 containing a zinc
finger gene cluster

JOURNAL Unpublished
2 (bases 1 to 176928)

REFERENCE Kodoyanni,V., Ge,Y., Krummel,G.K., Kvikstad,E., Grable,L.,
Severin,J., Gordon,L., Shannon,M., Brower,A., Olsen,A.S. and
Smith,L.M.

TITLE Direct Submission
JOURNAL Submitted (18-OCT-2000) Department of Chemistry, University of
Wisconsin, 1101 University Ave., Madison, WI 53706, USA
Map and sequence oriented from centromere to q-telomere. BC347040
(CTC-512J12) is adjacent to BC228680 (CTC-204F22, AC074311) on the
left, and overlaps BC101503 (CTB-188D12, AC069278) on the right
from bases 176,142 to 176,929 of this accession. Additional chr 19
map and sequence information are available at:
<http://www-bio.lnln.gov/genome/genome.html>.

FEATURES
Source
1. 176928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2"
/clone="CTC-512J12 (LLNL clone name BC347040)"
/clone_1lb="CIT-HSPC"
complement(<1..>414)
/gene="ZFP93"
/product="ZFP93 mRNA, partial CDS and 3'UTR"
complement(<1..>414)
/gene="ZFP93"
/note="zinc finger protein homologous to Zfp93 in mouse"
complement(104..>412)
/gene="ZFP93"
/note="zinc finger protein homologous to Zfp93 in mouse;
Complete human protein sequence not currently available.
Location refers to currently available ZFP93 sequence"
/codon_start=1
/product="ZFP93"
/protein_id="AAC23969.1"
/db_xref="GI:10864173"
/translation="SORSNLOVHIIHTGKPKCECKEFSWSAGLSAHORVHTGE
KPYTCOCGKGFSQASHFTHQVHTGERPYICDVCKGFSORSLSLHQRVHTGNTL
"
2778..2849
/rpt_family="L1M4"
complement(3326..3528)
/rpt_family="Alusg/x"
3830..3886
/rpt_family="A-rich"
3887..4262
/rpt_family="L1M4"
4348..4992
/rpt_family="L1M4"
4996..5138
/rpt_family="L1M4"
5133..6055
/rpt_family="L1PAL5-16"
6080..8755
/rpt_family="L1"
complement(8808..8888)
/rpt_family="L1P3"
8895..9023
/rpt_family="L1MD3"
9553..9961
/rpt_family="L1M4"
complement(10360..10650)
/rpt_family="Alusp"
complement(10654..10723)
/rpt_family="L2"
10724..11008
/rpt_family="Aluub"
complement(11009..11277)
/rpt_family="L2"

repeat_region complement(11278..11583)
/rpt_family="Alusx"
repeat_region complement(11584..11636)
/rpt_family="L2"
repeat_region 12104..12186
/rpt_family="MER112"
repeat_region 12353..12465
/rpt_family="MER20"
repeat_region 13026..13244
/rpt_family="MIR"
complement(13281..13445)
/rpt_family="L1ME"
repeat_region 13446..13747
/rpt_family="Alusx"
repeat_region complement(13748..13797)
/rpt_family="L1ME"
repeat_region complement(14360..14650)
/rpt_family="Alusx"
repeat_region 15265..15288
/rpt_family="A1-rich"
repeat_region 15409..15632
/rpt_family="Aluio"
repeat_region 15633..16079
/rpt_family="TA)n"
repeat_region complement(16373..16532)
/rpt_family="MER91A"
repeat_region 16634..16768
/rpt_family="MIR"
repeat_region complement(17464..17622)
/rpt_family="MIR"
repeat_region complement(18321..18758)
/rpt_family="MER110A"
repeat_region complement(18768..19594)
/rpt_family="LTR1"
19774..19927
/standard_name="NIB531"
/note="RH55859"
19776..19915
/standard_name="STSG40922"
/note="RH78239"
repeat_region complement(20076..20252)
/rpt_family="Trigger2a"
repeat_region complement(20253..20489)
/rpt_family="Alusp"
repeat_region complement(20490..20752)
/rpt_family="Trigger2a"
repeat_region complement(20780..21037)
/rpt_family="L1MC5"
repeat_region complement(21049..21454)
/rpt_family="LTR34"
repeat_region complement(21455..21986)
/rpt_family="LTR8"
repeat_region complement(21987..22292)
/rpt_family="Aluya5"
repeat_region complement(22293..22451)
/rpt_family="LTR8"
repeat_region complement(22452..22726)
/rpt_family="LTR34"
22938..23075
/rpt_family="MER77"
repeat_region 23099..23672
/rpt_family="MER77"
repeat_region complement(23732..24027)
/rpt_family="Aluio"
repeat_region complement(24056..24219)
/rpt_family="Trigger1"
repeat_region 24220..24506
/rpt_family="Aluy"
repeat_region complement(24507..25062)
/rpt_family="Trigger1"
repeat_region 25062..25121
/rpt_family="LTR16A"
repeat_region 25323..25386


```

* 7137 8692: contig of 1556 bp in length
* 8693 8712: gap of unknown length
* 8713 10229: contig of 1517 bp in length
* 10230 10249: gap of unknown length
* 10250 11190: contig of 941 bp in length
* 11191 11210: gap of unknown length
* 11211 12151: contig of 941 bp in length
* 12152 12171: gap of unknown length
* 12172 12930: contig of 759 bp in length
* 12931 12950: gap of unknown length
* 12951 14261: contig of 1311 bp in length
* 14262 14281: gap of unknown length
* 14282 15628: contig of 1347 bp in length
* 15629 15648: gap of unknown length
* 15649 16920: contig of 1272 bp in length
* 16921 16940: gap of unknown length
* 16941 18079: contig of 1139 bp in length
* 18080 18099: gap of unknown length
* 18100 18946: contig of 847 bp in length
* 18947 18966: gap of unknown length
* 18967 19946: contig of 980 bp in length
* 19947 19966: gap of unknown length
* 19967 20787: contig of 821 bp in length
* 20788 20807: gap of unknown length
* 20808 22768: contig of 1961 bp in length
* 22769 22788: gap of unknown length
* 22789 23673: contig of 885 bp in length
* 23674 23693: gap of unknown length
* 23694 25232: contig of 1539 bp in length
* 25233 25252: gap of unknown length
* 25253 27310: contig of 2058 bp in length
* 27311 27330: gap of unknown length
* 27331 28319: contig of 989 bp in length
* 28320 28339: gap of unknown length
* 28340 29740: contig of 1401 bp in length
* 29741 29760: gap of unknown length
* 29761 31375: contig of 1615 bp in length
* 31376 31395: gap of unknown length
* 31396 33359: contig of 1964 bp in length
* 33360 33379: gap of unknown length
* 33380 34897: contig of 1518 bp in length
* 34898 34917: gap of unknown length
* 34918 36591: contig of 1674 bp in length
* 36592 36611: gap of unknown length
* 36612 37630: contig of 1019 bp in length
* 37631 37650: gap of unknown length
* 37651 38842: contig of 1192 bp in length
* 38843 38862: gap of unknown length
* 38863 40413: contig of 1551 bp in length
* 40414 40433: gap of unknown length
* 40434 42428: contig of 1995 bp in length
* 42429 42448: gap of unknown length
* 42449 43991: contig of 1543 bp in length
* 43992 44011: gap of unknown length
* 44012 45215: contig of 1204 bp in length
* 45216 45235: gap of unknown length
* 45236 47171: contig of 1936 bp in length
* 47172 47191: gap of unknown length
* 47192 49075: contig of 1884 bp in length
* 49076 49095: gap of unknown length
* 49096 51980: contig of 2885 bp in length
* 51981 52000: gap of unknown length
* 52001 53551: contig of 1551 bp in length
* 53552 53571: gap of unknown length
* 53572 56030: contig of 2459 bp in length
* 56031 56050: gap of unknown length
* 56051 57435: contig of 1385 bp in length
* 57436 57455: gap of unknown length
* 57456 59333: contig of 1878 bp in length
* 59334 59353: gap of unknown length
* 59354 61391: contig of 2038 bp in length
* 61392 61411: gap of unknown length
* 61412 64200: contig of 2789 bp in length

```

```

* 64201 64220: gap of unknown length
* 64221 66723: contig of 2503 bp in length
* 66724 66743: gap of unknown length
* 66744 68487: contig of 1744 bp in length
* 68488 68507: gap of unknown length
* 68508 70631: contig of 2124 bp in length
* 70632 70651: gap of unknown length
* 70652 73101: contig of 2450 bp in length
* 73102 73121: gap of unknown length
* 73122 76993: contig of 3872 bp in length
* 76994 77013: gap of unknown length
* 77014 80128: contig of 3115 bp in length
* 80129 80148: gap of unknown length
* 80149 81844: contig of 1696 bp in length
* 81845 81864: gap of unknown length
* 81865 83437: contig of 1573 bp in length
* 83438 83457: gap of unknown length
* 83458 86257: contig of 2800 bp in length
* 86258 86277: gap of unknown length
* 86278 89119: contig of 2842 bp in length
* 89120 89139: gap of unknown length
* 89140 91732: contig of 2593 bp in length
* 91733 91752: gap of unknown length
* 91753 94902: contig of 3150 bp in length
* 94903 94922: gap of unknown length
* 94923 98773: contig of 3851 bp in length
* 98774 98793: gap of unknown length
* 98794 101186: contig of 2393 bp in length
* 101187 101206: gap of unknown length
* 101207 104287: contig of 3081 bp in length
* 104288 104307: gap of unknown length
* 104308 108276: contig of 3969 bp in length
* 108277 108296: gap of unknown length
* 108297 112998: contig of 4702 bp in length
* 112999 113018: gap of unknown length
* 113019 120694: contig of 7676 bp in length
* 113019 120695: gap of unknown length
* 120695 127015: gap of unknown length
* 127015 127047: contig of 6332 bp in length
* 127047 127066: gap of unknown length
* 127067 138938: contig of 11872 bp in length.

```

```

FEATURES
Source
Location/Qualifiers
1..138938
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/cclone="RPCI98-9B18"

```

```

BASE COUNT 38349 a 29399 c 29584 g 40293 t 1313 others
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 2; Length 138938;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 ttctgcatctgctgttggaagcg 25
Db 134576 TTGCGGATGTCTCTTTGAAGCG 134553

```

```

RESULT 28
AC084239/c DNA PRI 18-Oct-2000
LOCUS Homo sapiens chromosome 19, BAC CMC-512J12 (BC347040), complete
DEFINITION
sequence.
AC084239
AC084239.1 GI:10864171
VERSION
AC084239
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens

```

```

REFERENCE
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 176928)
Kodoyanagi, V., Ge, Y., Krummel, G. K., Kvistad, E., Grable, L.,
Severin, J., Gordon, L., Shannon, M., Brower, A., Olsen, A. S. and

```

```

misc_feature      complement(21990..22126)
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 2, quality: good, score: 52.000"
repeat_region     /rpt_family="ALU"
                  complement(22667..22933)
repeat_region     /rpt_family="ALU"
                  complement(23080..23300)
repeat_region     /rpt_family="ALU"
                  complement(24284..24629)
misc_feature      /rpt_family="THE1"
                  24753..24861
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 0, quality: good, score: 58.000"
repeat_region     /rpt_family="ALU"
                  25445..25663
repeat_region     /rpt_family="THE1"
                  complement(25693..26047)
repeat_region     /rpt_family="ALU"
                  26097..26382
repeat_region     /rpt_family="ALU"
                  complement(26593..26981)
misc_feature      /rpt_family="L1"
                  27076..27220
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 0, quality: excellent, score: 100.000"
repeat_region     /rpt_family="THE1"
                  complement(27316..27383)
misc_feature      /rpt_family="MLT1"
                  28274..28399
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 1, quality: good, score: 62.000"
misc_feature      28800..29086
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 2, quality: excellent, score: 88.000"
misc_feature      29251..31136
                  /note="BLASTX similarity to finger protein 2, placental -
                  human; ZINC FINGER PROTEIN ZFP-38 (CTF1N51) (TRANSCRIPTION
                  FACTOR RFX49) >gnl|P101955 (D10630) zinc finger
                  protein [Mus musculus]
                  EST similarities: 4233166, T12758"
repeat_region     32217..32550
                  /rpt_family="THE1"
misc_feature      33424..33521
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 0, quality: excellent, score: 100.000"
repeat_region     /rpt_family="MLT1"
                  complement(33811..33919)
repeat_region     35479..38000
                  /rpt_family="L1"
misc_feature      35672..35800
                  /rpt_family="L1"
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 1, quality: marginal, score: 46.000"
misc_feature      36135..36344
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 2, quality: good, score: 63.000"

BASE COUNT      10578 a 8454 c 7972 g 10996 t
ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 38000;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 gttgctgatactcgtctgtgaac 24
DB 11472 GTTGCTGGCTCTGCTTTTGTATGC 11495

RESULT 27
AC010111 138938 bp DNA HTG 18-FEB-2000
LOCUS     Drosophila melanogaster clone RPC198-9B18, *** SEQUENCING IN
DEFINITION
PROGRESS ***, 63 unordered pieces.
ACCESSION AC010111
VERSION   AC010111.4 GI:6996750
KEYWORDS  HTG; HTGS_PHASE1.

```

```

SOURCE
ORGANISM
REFERENCE
AUTHORS
COMMENT
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
Direct Submission
Unpublished
2 (bases 1 to 138938)
Worley,K.C.
Direct Submission
Submitted (13-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 18, 2000 this sequence version replaced gi:5902227.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: BRDY
Center clone name: RPC198-9B18
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 29% of reads
Chemistry: Dye-terminator Big Dye: 10% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 84003 bases at least Q40
Consensus quality: 116061 bases at least Q20
Estimated insert size: 128604; sum-of-coverage estimation
Quality coverage: 1x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1114: contig of 1114 bp in length
* 1115 1134: gap of unknown length
* 1135 2041: contig of 907 bp in length
* 2042 2061: gap of unknown length
* 2062 3598: contig of 1537 bp in length
* 3599 3618: gap of unknown length
* 3619 4998: contig of 1380 bp in length
* 4999 5018: gap of unknown length
* 5019 5989: contig of 971 bp in length
* 5990 6010: gap of unknown length
* 6011 7116: contig of 1107 bp in length
* 7117 7136: gap of unknown length

```

```

* 242193 245672: contig of 3480 bp in length
* 245673 245772: gap of unknown length
* 245773 247971: contig of 2199 bp in length
* 247972 248071: gap of unknown length
* 248072 250703: contig of 2632 bp in length
* 250704 250803: gap of unknown length
* 250804 252676: contig of 1873 bp in length
* 252677 252776: gap of unknown length
* 252777 255037: contig of 2261 bp in length
* 255038 255137: gap of unknown length

Query Match
Best Local Similarity 79.28; Score 19.8; DB 2; Length 303717;
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtctgcatctcgtcttgaag 23
||||| ||||| ||||| |||||
Db 231708 GTTCTGCACTGCTGTGAAG 231730

RESULT 25
AC014347/c AC014347 22329 bp DNA HTG 16-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***; in ordered
DEFINITION pieces.
ACCESSION AC014347
VERSION AC014347.1 GI:6436988
KEYWORDS HTG: HTGS-PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 22329)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212385 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..22329
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 6272 a 5051 c 4780 g 6226 t

ORIGIN
6272 a 5051 c 4780 g 6226 t

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 2; Length 22329;
Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 ttgctgcatctcgtcttgaagc 25
||||| ||||| ||||| |||||
Db 1605 TTGCGGATGCTGTGAAGCG 1582

RESULT 26
AF025422 AF025422 38000 bp DNA PRI 20-OCT-1997
LOCUS Homo sapiens chromosome 19 cosmid F15386, genomic sequence,
DEFINITION complete sequence.
ACCESSION AF025422
VERSION AF025422.1 GI:2547408
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 38000)
AUTHORS Rank, D.R., Westphall, M.S., Ono, T., Berggren, W.T., Lamerdin, J.,
Ashworth, L. and Smith, L.M.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1997) Department of Chemistry, University of
Wisconsin-Madison, 1101 University Ave, Madison, WI 53706, USA
FEATURES
source
1..38000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="F15386"
/chromosome="19"
/map="19q13.2"
/cell_line="UV5HL9-5B"
/clone_lib="L19NC02 F2 chromosome 19-specific cosmid
library"
/note="cosmid library constructed at LNL from flow-sorted
chromosomes from hybrid UV5HL9-5B, which carries
chromosome 19 as its only human chromosome."
273..435
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 53.000"
1322..1447
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 60.000"
1456..1585
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: marginal, score: 48.000"
3163..3198
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 55.000"
3296..3495
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 92.000"
4818..5102
/rpt_family="L1"
5368..5528
/rpt_family="MER21"
5629..5672
/rpt_family="MER21"
6262..6374
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 74.000"
7274..7318
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 79.000"
7733..7933
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 57.000"
8100..8359
/rpt_family="ALU"
complement(8787..9087)
/rpt_family="ALU"
complement(10868..11032)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: marginal, score: 47.000"
11524..11677
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 56.000"
complement(14364..14556)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 65.000"
complement(16428..16579)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 76.000"
18436..18715
/rpt_family="ALU"
19710..19989
/rpt_family="ALU"
20245..20696
/rpt_family="MER21"
complement(21314..22104)
/rpt_family="L1"

```

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Wellington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
 and Gibbs, R.
 Direct Submission
 2 (bases 1 to 303717)
 Morley, K.C.
 Submitted (22-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 29, 2000 this sequence version replaced gi:1141926.

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: HAYW
 Center clone name: RP11-51AF2
 Summary Statistics
 Sequencing Vector: M13: L08821
 Chemistry: Dye-Primer Bodypy: 3% of reads
 Chemistry: Dye-terminator Big Dye: 97% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 221091 bases at least Q40
 Consensus quality: 252136 bases at least Q40
 Consensus quality: 268230 bases at least Q20
 Estimated insert size: 265010; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 NOTE: This sequence may represent more than one clone.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 86 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 9894: contig of 9894 bp in length
 * 9895 9994: gap of unknown length
 * 9995 20133: contig of 10139 bp in length
 * 20134 20233: gap of unknown length
 * 20234 29489: contig of 9256 bp in length
 * 29490 29589: gap of unknown length
 * 29590 36626: contig of 7037 bp in length
 * 36627 36726: gap of unknown length
 * 36727 42239: contig of 5513 bp in length
 * 42240 42339: gap of unknown length
 * 42340 50165: contig of 7826 bp in length
 * 50166 50265: gap of unknown length
 * 50266 58725: contig of 8460 bp in length
 * 58726 58825: gap of unknown length
 * 58826 65134: contig of 6309 bp in length
 * 65135 65234: gap of unknown length
 * 65235 71642: contig of 6408 bp in length
 * 71643 71742: gap of unknown length
 * 71743 77153: contig of 5411 bp in length
 * 77154 77253: gap of unknown length
 * 77254 84462: contig of 7209 bp in length
 * 84463 84562: gap of unknown length
 * 84563 91495: contig of 6933 bp in length
 * 91496 91595: gap of unknown length
 * 91596 97457: contig of 5862 bp in length
 * 97458 97557: gap of unknown length
 * 97558 105030: contig of 7473 bp in length

105031 105130: gap of unknown length
 * 105131 110324: contig of 5194 bp in length
 * 110325 110424: gap of unknown length
 * 110425 115824: contig of 5400 bp in length
 * 115825 115924: gap of unknown length
 * 115925 121202: contig of 5278 bp in length
 * 121203 121302: gap of unknown length
 * 121303 124865: contig of 3563 bp in length
 * 124866 131022: contig of 6057 bp in length
 * 131023 131122: gap of unknown length
 * 131123 134925: contig of 3803 bp in length
 * 134926 135025: gap of unknown length
 * 135026 139918: contig of 4892 bp in length
 * 139918 140018: gap of unknown length
 * 140018 144220: contig of 4203 bp in length
 * 144221 144320: gap of unknown length
 * 144321 147439: contig of 3119 bp in length
 * 147440 147539: gap of unknown length
 * 147540 152505: contig of 4966 bp in length
 * 152506 152605: gap of unknown length
 * 152606 156006: contig of 3401 bp in length
 * 156007 156106: gap of unknown length
 * 156107 161617: contig of 5511 bp in length
 * 161618 161717: gap of unknown length
 * 161718 167424: contig of 5707 bp in length
 * 167425 167524: gap of unknown length
 * 167525 171814: contig of 4290 bp in length
 * 171815 171914: gap of unknown length
 * 171915 174719: contig of 2805 bp in length
 * 174720 179373: gap of unknown length
 * 179374 179473: contig of 4554 bp in length
 * 179474 183064: contig of 3591 bp in length
 * 183065 183164: gap of unknown length
 * 183165 186280: contig of 3116 bp in length
 * 186281 186380: gap of unknown length
 * 186381 190385: contig of 4005 bp in length
 * 190386 190485: gap of unknown length
 * 190486 193709: contig of 3224 bp in length
 * 193710 193809: gap of unknown length
 * 193810 197580: contig of 3771 bp in length
 * 197581 197680: gap of unknown length
 * 197681 200912: contig of 3232 bp in length
 * 200913 201012: gap of unknown length
 * 201013 204830: contig of 3818 bp in length
 * 204831 204930: gap of unknown length
 * 204931 208131: contig of 3201 bp in length
 * 208132 210922: contig of 2691 bp in length
 * 210923 211022: gap of unknown length
 * 211023 214347: contig of 3325 bp in length
 * 214348 214447: gap of unknown length
 * 214448 216949: contig of 2502 bp in length
 * 216950 217049: gap of unknown length
 * 217050 219757: contig of 2708 bp in length
 * 219758 219857: gap of unknown length
 * 219858 223103: contig of 3246 bp in length
 * 223104 223203: gap of unknown length
 * 223204 226009: contig of 2806 bp in length
 * 226010 226109: gap of unknown length
 * 226110 229109: contig of 3000 bp in length
 * 229109 232471: gap of unknown length
 * 232472 232571: gap of unknown length
 * 232572 232684: contig of 2713 bp in length
 * 232685 235384: gap of unknown length
 * 235385 238653: contig of 3269 bp in length
 * 238654 238753: gap of unknown length
 * 238754 240129: contig of 1376 bp in length
 * 240130 240229: gap of unknown length
 * 240230 242092: contig of 1863 bp in length
 * 242093 242192: gap of unknown length

```

* 89798 94730: contig of 4933 bp in length
* 94731 94830: gap of unknown length
* 94831 100596: contig of 5766 bp in length
* 100597 100696: gap of unknown length
* 100697 105812: contig of 5116 bp in length
* 105813 105913: gap of unknown length
* 105913 113093: contig of 7181 bp in length
* 113094 113193: gap of unknown length
* 113194 118164: contig of 4971 bp in length
* 118165 118264: gap of unknown length
* 118265 125193: contig of 6929 bp in length
* 125194 125293: gap of unknown length
* 125294 130564: contig of 5271 bp in length
* 130565 130664: gap of unknown length
* 130665 136299: contig of 5635 bp in length
* 136300 136399: gap of unknown length
* 136400 142500: contig of 6101 bp in length
* 142501 142600: gap of unknown length
* 142601 146623: contig of 4023 bp in length
* 146624 146723: gap of unknown length
* 146724 150634: contig of 3911 bp in length
* 150635 150734: gap of unknown length
* 150735 157354: contig of 6620 bp in length
* 157355 157454: gap of unknown length
* 157455 161824: contig of 4370 bp in length
* 161825 161924: gap of unknown length
* 161925 166911: contig of 4987 bp in length
* 166912 167011: gap of unknown length
* 167012 171639: contig of 4628 bp in length
* 171640 171740: gap of unknown length
* 171741 178016: contig of 6277 bp in length
* 178017 178116: gap of unknown length
* 178117 184085: contig of 5969 bp in length
* 184086 184185: gap of unknown length
* 184186 189226: contig of 5041 bp in length
* 189227 189326: gap of unknown length
* 189327 194366: contig of 5040 bp in length
* 194367 194466: gap of unknown length
* 194467 198530: contig of 4064 bp in length
* 198531 198630: gap of unknown length
* 198631 203086: contig of 4456 bp in length
* 203087 203186: gap of unknown length
* 203187 206934: contig of 3748 bp in length
* 206935 207034: gap of unknown length
* 207035 211278: contig of 4244 bp in length
* 211279 211378: gap of unknown length
* 211379 214720: contig of 3342 bp in length
* 214721 214820: gap of unknown length
* 214821 217723: contig of 2903 bp in length
* 217724 217823: gap of unknown length
* 217824 221500: contig of 3677 bp in length
* 221501 221600: gap of unknown length
* 221601 225644: contig of 4044 bp in length
* 225645 225744: gap of unknown length
* 225745 229281: contig of 3537 bp in length
* 229282 229381: gap of unknown length
* 229382 232302: contig of 2921 bp in length
* 232303 232402: gap of unknown length
* 232403 235101: contig of 2699 bp in length
* 235102 235201: gap of unknown length
* 235202 237498: contig of 2297 bp in length
* 237499 237598: gap of unknown length
* 237599 240299: contig of 2701 bp in length
* 240300 240399: gap of unknown length
* 240400 242831: contig of 2432 bp in length
* 242832 242931: gap of unknown length
* 242932 246173: contig of 3242 bp in length
* 246174 246273: gap of unknown length
* 246274 248692: contig of 2419 bp in length
* 248693 251172: gap of unknown length
* 248793 251172: contig of 2386 bp in length
* 251173 251272: gap of unknown length
* 251273 254217: contig of 2945 bp in length

```

```

Query Match          79.2%  Score 19.8;  DB 2;  Length 302558;
Best Local Similarity 91.3%  Pred. No. 1.5e+02;
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 gttcgtgacgtcgttgaag 23
Db 192545 GTTCTGCAACTGCTTGAAG 192567
|||||

```

```

RESULT 24
AC063930
LOCUS Homo sapiens chromosome 3 clone RP11-514F2, *** SEQUENCING IN
DEFINITION
ACCESSION AC063930
VERSION AC063930.17 GI:11995500
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Chordata: Vertebrata: Euteleostomi:
Eukaryota: Euteleostomi: Primates: Catarrhini: Hominoidea: Homo.
Mammalia: Euteleostomi: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 303717)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Alshrocks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Blinag,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowle,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burrell,C., Burrell,K.U., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Demm,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flegg,N., Ford,J.,
Foster,P., Frintz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtenarge,O., Ilean,C., Liu,J., Liu,W.,
Louiaged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Kapur,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Monabadi,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockenkro,S.,
Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Peery,J., Ren,Y., Rives,M., Rojas,A., Rojokokan,I., Rolfe,M.,
Ruliz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,A.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

```



```

/chrnslation="DCXLRTRXGEXNLVSCIALGLPLPIASGTMMHEAPGAVNNGTE
XVVDODRIFISOLTVEENTLHTAOMAKGIAGSISDSYDFEPELEYLPOLPASCINPI
OOYOLALANALVTSPRLLIDEPKSGTQULTYVLAOLILFNPRLGILSVLLAECHLS
LIRRYADRFCLLIGRNVAQGRVSLDDPLVMMWPAASFO"
BASE COUNT      1512 a      2642 c      2605 g      1453 t      15 others
ORIGIN

```

COMMENT

Worley, K.C.
Direct Submission
Submitted (09-Oct-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 5, 2001 this sequence version replaced gi:12007667.

Query Match	79.28;	Score 19.8;	DB 1;	Length 8227;
Best Local Similarity	91.38;	Pred. No. 1e+02;		
Matches 21; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Qy	3	tgcgtgacatcgtctgtttgaagcg	25
Db	4087	TGCTTGAACACTGCTGTTTGAAGCG	4109

RESULT	21
AC084015/c	
LOCUS	
DEFINITION	HMG 07-JAN-2005
	RPL1-995B13, WORKING DRAFT
AC084015	185738 bp DNA
	Homo sapiens chromosome 10 clone RPL1-995B13,
SEQUENCE	23 unordered pieces.
AC084015	
AC084015.9	GI:12039118
KEYWORDS	HMG; HUGS_PHASE1; HUGS_DRAFT.
SOURCE	human.

REFERENCE
AUTHORS

TITLE	Direct Submission
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 185738)

Worley, K.C.
 Direct Submission
 Submitted (09-Oct-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jan 5, 2001 this sequence version replaced gi:12007667.

 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: HHOB
 Center clone name: RP11-995B13

 Summary Statistics
 Sequencing vector: M13, L08821
 Chemistry: Dye-terminator Big Dye: 46% of reads
 Assembly: Dye-terminator Big Dye: 54% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 166584 bases at least Q40
 Consensus quality: 174801 bases at least Q30
 Consensus quality: 179020 bases at least Q20
 Estimated insert size: 179024: sum-of-coverage estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 3.6x in Q20 bases; sum-of-coverage estimation

 NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 NOTE: This is a 'working draft' sequence. It currently
 consists of 23 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1	44236:	contig of 44236 bp in length
44237	44336:	gap of unknown length
44337	63479:	contig of 19143 bp in length
63480	63579:	gap of unknown length
63580	81827:	contig of 18248 bp in length
81828	81927:	gap of unknown length
81928	93064:	contig of 11137 bp in length
93065	93164:	gap of unknown length
93165	105462:	contig of 12298 bp in length
105463	105562:	gap of unknown length
105563	115530:	contig of 9968 bp in length
115531	115630:	gap of unknown length
115631	124121:	contig of 8491 bp in length
124122	124221:	gap of unknown length
124222	133919:	contig of 9698 bp in length
133920	134019:	gap of unknown length
134020	141345:	contig of 7326 bp in length
141346	141445:	gap of unknown length
141446	147781:	contig of 6336 bp in length
147782	147881:	gap of unknown length
147882	153389:	contig of 5508 bp in length
153390	153489:	gap of unknown length
153490	158140:	contig of 4651 bp in length
158141	158240:	gap of unknown length
158241	163014:	contig of 4674 bp in length
163015	163014:	gap of unknown length
163015	166903:	contig of 3889 bp in length
166904	167003:	gap of unknown length
167004	170364:	contig of 3361 bp in length
170365	170464:	gap of unknown length
170465	173923:	contig of 3459 bp in length
173924	174023:	gap of unknown length
174024	175899:	contig of 1876 bp in length
175900	175999:	gap of unknown length
176000	178095:	contig of 2096 bp in length
178096	178195:	gap of unknown length
178196	180140:	contig of 1945 bp in length

REFERENCE 2 (bases 1 to 8227)
 AUTHORS Entcheva, P., Liebl, W. and Streil, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES
 source 1..8227
 /organism="uncultured bacterium pCosa1"
 /db_xref="taxon:143793"
 /clone="pCosa1"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 /gene="hutu"
 <1..502
 /gene="hutu"
 /note="putative urocanate hydratase"
 /codon_start=2
 /transl_table=11
 /product="urocanase-like protein"
 /protein_id="AAG53593.1"
 /db_xref="GI:12407618"
 /translation="PRERISFQGLPARICWGLEMRKGLGLAENMVRGEVAPYI
 GRHDGSSVASPNRETFAMODGSDAYSDFLLNALMTAGATWVSLHGGVGWGF
 SOHAGMYVCQGTDDAARLARVLRHNDPATGVRHADAGYDAIDACEGLPMIA
 ATQGOH"
 502..2031
 /gene="huth"
 502..2031
 /gene="huth"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="histidine ammonia-lyase-like protein"
 /protein_id="AAG53586.1"
 /db_xref="GI:12407611"
 /translation="MNALITPFGTLTLAOLROWOPLQLTDESHEAINDVACVE
 AIVAGRTAVGINTFGFLAOTRIATHDLELQSLVLSHAAGVEPLDDIYRLMW
 LKINSLARGEIGRLSVYQIALVNAVGVSDPAKSGASGLAPLALMSLTLLGE
 GKARYRGEMLPAAATLQKAGLAPVTLAAKEGLALNGTOASTALRGLEFEDLEFS
 AVCGALITTEAVLIGSRPFDAKRIHEVGRGOQIDAALFRHVLTDTSAIESHHNCDK
 VQDYSLRCOPQWAGCTLOROVAFVLIVESNAVSPNPLVPAENEMVTRGNHAP
 VAMADNLALAIETGALSERRIALMMDKHSOLPPLVTRGVNSGFMIAQVTAAL
 ASENKGCCHPTTTSVDKPCRPANQEDHVSMAAPAGRRIMEMGNTRGVLAEMLAAC
 QGADLRGLTSSPLLEQARQSCGEVAVHYDDRFAPDIEAALSLNKGSLVGLPAF
 L"
 complement(2130..2606)
 /note="ORF1: similar to hypothetical 17.1 kDa protein in
 mod-bioa intergenic region"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAG53587.1"
 /db_xref="GI:12407612"
 /translation="MKLISODLRDGLPLRHVFNMGYEGDNI SPLHAMDVEPSGK
 SPVYTCYDPADPTDSCGMHWIVANITPADTVLQSGSSSI VALPEGAVERTDFFGAG
 YCGAAPKGETHRIITVHALNVERLEIVDEASGAMGVNHFHTLGSASTAMYS"
 complement(2651..3940)
 /gene="bioa"
 complement(2651..3940)
 /gene="bioa"
 /note="7.8-diamino-pelargonic acid aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA aminotransferase Bioa"
 /protein_id="AAG53588.1"
 /db_xref="GI:12407613"
 /translation="MTTDDLPDRHWHIPYTSMPOLPYVVEAHGCELTIASGEQ
 LVDGSSMAAIIHGYNHPRLNAKMOIDRMSVVMGGITTHPAVDLCRLVAMTPE
 LECVFPADSSVAVYAKKALDIYHAKGESRORTTFTRNGYHGFDFGKAVSCDDPNS
 MSLMKGYLPENLEFAPPOSFDDGAMDEMVGVLAPLMAHRIEIAVILIEPIVQAG
 GKRRYHPEWLKRIKMCQDREGILLIADIELATGFGTKLFAECENAGITPDICTICGAL
 TGCTMTMSATLTTROVAERTISNGACCFMHPFTFGCNPLACAVASLSALLSGEMQD

gene
 CDS
 4029..5069
 /gene="biob"
 4029..5069
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase Biob"
 /protein_id="AAG53589.1"
 /db_xref="GI:12407614"
 /translation="MAHQARWTLTQVTLAEKPILELLFEAQOIHROHFPQOIGVST
 LLSIKTGACPEDECKKCPQASRYKGTGLERLMEVBYLESAAQAKNAGSTRCKGAAM
 KNHEHMRMYLQEMVQGVKAMGLQACMTIGTLDHQAQLASAGLDYTNHNLDTPEF
 YGNIITTRTYQERLDTLDKRVKAGIKVSGGIVGICETVTDRAGLLOLANLPPEPS
 VPINMLVKYKGTPLADNDVXAFDXIRTSVAVXAKAPTYLIVLSTGREQMOTQAMC
 FMAGANSIFYGCRLTTPNPEEDKDVQLFRKLGILNPQAVAVLGDNEQHQLEQULN
 ADTDOFYNNAVTV"
 5066..6223
 /gene="bioc"
 5066..6223
 /gene="bioc"
 /note="7-keto-8-amino-pelargonic acid synthetase; similar
 to BiocF"
 /codon_start=1
 /transl_table=11
 /product="7-keto synthetase"
 /protein_id="AAG53590.1"
 /db_xref="GI:12407615"
 /translation="MTMOORIDAAIDARRAADALRVYPVENGAGRPLIREORRYCNF
 SGNDYLGLOSHPELIRAMQOGAEOYGVSGSGSGGYTHAHOALEHLAGLQTPRA
 LFIISGFANNOALIAITDKEDRIYADRISHSLEEAASHPAQLRRRHANDVQDS
 LLDKPGHGOQLVTEGTFMSDQSDAPLGVITAAORQSGMLVDAHSGIVTGPBGRG
 SAAQGVKPELLVTFEGKGVSGAIVLCSDVADYLIQFAHLLYSTMSPEQAVAL
 LAEFVIRSDENARRQRLAALIDQFRAVGRPLPQITASQSIQPLIVGENARALHL
 AQRLEKRGIMVTAIRPPVPAGTARLRILITAAHQPADITALEALHAGE"
 6207..6962
 /gene="bioc"
 6207..6962
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis Bioc-like protein"
 /protein_id="AAG53591.1"
 /db_xref="GI:12407616"
 /translation="MSQYKQALIAAFGBAOSYSRRHALORQSDGLRALLAENTFA
 SYLDGCGGSGNSLAMRQVSVTAMDLCAPMLEBARRNOADSYLQADIEAIPLADA
 OFSLAMSHLAVOXCASLPRALAEIWRVTLQGGKVAFTLLNGSLPELNQAMQAVDTP
 HANRPLSGOQVNDALSGMQYRCAYOVVTLTFDDALSAMHSLKIGATHTLHNGRAAKPL
 TRGOLQOLAQAMPQOQAKFPLSYQLFHGIIRD"
 6940..7680
 /gene="bioc"
 6940..7680
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /product="dehydrobiotin synthase Bioc"
 /protein_id="AAG53592.1"
 /db_xref="GI:12407617"
 /translation="WELSNVTERYFVYGTDTVEGKTYASTALLQAAALGRHAGYKRP
 VASGSEMTAOGILRNSDALALQNSLSLASYAVNVPYFAEPTSTHISADSORPIEFA
 TLSAGILRALETDQADVILVEGAGGWTPLSGOTFADWVMAERLPIIVLVGKACGINH
 AMLTAQAVIRQAGLHFRAGWAGVYVPGKRHAHEVMATLRRVLSAPLLEIPIWLGQADT
 SAVGQYLDITLALTPSAPSSGOSAGHPAR"
 complement(7619..>8227)
 /gene="elisa"
 complement(7619..>8227)
 /gene="elisa"
 /note="putative urea/short-chain amide ABC transporter"
 /codon_start=1
 /transl_table=11
 /product="ABC transporter-like protein"
 /protein_id="AAG53594.1"

AUTHORS Sakurai, N.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry: 2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan (E-mail: oec0101@niftyserve.or.jp, nsakurai@dbj.nig.ac.jp, Tel: 048-433-2545, Fax: 048-433-2540)
REFERENCE 2 (bases 1 to 7215)
AUTHORS Sakurai, N., Imai, Y., Akatsuka, H., Kawai, E., Komatsubara, S. and Tosa, T.
TITLE Complete nucleotide sequence of biotin operon of *Serratia marcescens*
JOURNAL Unpublished (1993)
COMMENT Submitted (25-Aug-1993) to DDBJ by: Naoki Sakurai
 Res. Lab. of Applied Biochemistry
 Tanabe Seiyaku Co., Ltd.
 2-50 Kawagishi-2-chome
 Toda, Saitama 335
 Japan
 Phone: 048-433-2545
 Email: nsakurai@dbj.nig.ac.jp
 Fax: 048-433-2540.
FEATURES Location/Qualifiers
source 1..7215
 /organism="Serratia marcescens"
 /strain="Srt41"
 /db_xref="taxon:615"
 complement(1353..2630)
 /gene="bioc"
 complement(1353..2630)
 /gene="bioc"
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=1
 /product="7,8-diamino-pelargonic acid aminotransferase"
 (DAPA aminotransferase)
 /protein_id="BA04284.1"
 /db_xref="GI:402531"
 /translation="MSVTASDLAEFDRIHHPYTSMSRPLCYPIESASGVELQIADG
 RSLVDGMSWMAIIGHYHNPRLNOASQLEKSHVMFGGITHPAAISLCRLVEMP
 EALQCVFLADSGVAVEVSLKMALOYQAGGERQRIILTRHGCHGDTFGAMSCDPD
 NSMSTLYOGYLAFLPFAFAPQCRDEMRDIAPEFALLEQHAQGEVAAVILPVOG
 AGCMRTYHPYIKRVRVVAIKILLIADEIATGEGRTGKIFACEHAOVVPLICG
 KALTGYTILSATITTTTHVAETISNGAAGCFMHPGTPMGPLACAVDAITALAENR
 WQAVSAIETQLKRELLPLGRCRWPRVPLGAVGEKRPDVDAELQNGYVERGW
 IRPGKLIYDMPYPIEAQLSRLTAVAADAR"
 2716..3756
 /gene="biob"
 2716..3756
 /gene="biob"
 /codon_start=1
 /transl_table=1
 /product="biotin synthetase"
 /protein_id="BA04285.1"
 /db_xref="GI:402532"
 /translation="MMADRIHWGQAOALEDKPLLELLFEAGTVHROHFDPPQVYS
 TLLSTKACPEDECKYCPSPRYKTGLESERLQVEVLESARKAKANGSRCEMGVA
 WKNHERDMPYLQOVOCVRKAMGMECTKTLGTLGTAERLAEGCLDYNNINIDTSPF
 FVGSITTRSYOERLDITLDKVRAGIKVCSGIVGLGETVHADRAGLLVOLANPKPIH
 SVPIINMLVKGKPLADNDVDPDFIRTTIAVARIKMPSVYRSAGREONECTOM
 CFMAGANSIFYGCKLITTPNEDEKDLQLEFRKLLINQQTATENGDNQOOQVLAQOLL
 NADTAEFYNAP"
 3756..4904
 /gene="bioc"
 3756..4904
 /gene="bioc"
 /codon_start=1
 /transl_table=1
 /product="7-keto-8-aminopelargonic acid synthetase (KAPA
 synthetase)"
 /protein_id="BA04286.1"
 /db_xref="GI:402533"

/translation="MSWQORIEQALAEERLNAAYRRRTTEGGNGROIIRGDFLIYNF
 SGNDYLGISODARVIAAQQGARYGVNGSGHVTGSAHQALEQALAWLIGYPRAL
 LEFISYANQAVIALAMOKGDRILADRLSHASLILEAAGVAGRAPAPVAPQSTAGIARIC
 LKPCDGRALVTEGLSMQDGPALDELHRLTFAAGAMLVDDAHGIVRGQGRS
 CWQGVRELLIATPFGCAFGYSAAYACDEATATYLLQFRLIYTPAMPACALQ
 ALAKIRGDDLRPLDQNTTRFPGCAAPALATLTBDDTAIQPLVGDNRALDIATR
 LRKGLWVSAIRPLVPPGARGRLRLILAAHOSODIRLLLEVLDVNSQ"
 4888..5655
 /gene="bioc"
 4888..5655
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04287.1"
 /db_xref="GI:402534"
 /translation="MFSANDTVNKKQAVASAFSRAAGSDAAALORVGERLLMGSS
 HPGEDLDAGCGTGYFSRMREKGRKRYALDLACMDVARORAAHHYLLIGIEVP
 LPDAAMDICFSSLVVWCSDLPALAEILRYTRGCIYILSTLAGSLQELGAMQV
 DGERIVNAFLPLQIRTACAAVRIELVTELTLPVDMTLMRSKLGATHLHQRE
 GGIMSRGRIALQAAYPCROGFPLSYHLAVGYIRE"
 5648..6331
 /gene="biob"
 5648..6331
 /gene="biob"
 /function="conversion from 7,8-diaminopelargonic acid into
 dehydrobiotin"
 /codon_start=1
 /transl_table=1
 /product="dehydrobiotin synthetase (DTB synthetase)"
 /protein_id="BA04288.1"
 /db_xref="GI:912453"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 6331..7215
 /gene="bioc"
 6331..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04289.1"
 /db_xref="GI:912454"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04290.1"
 /db_xref="GI:912455"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04291.1"
 /db_xref="GI:912456"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04292.1"
 /db_xref="GI:912457"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04293.1"
 /db_xref="GI:912458"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04294.1"
 /db_xref="GI:912459"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04295.1"
 /db_xref="GI:912460"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04296.1"
 /db_xref="GI:912461"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04297.1"
 /db_xref="GI:912462"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04298.1"
 /db_xref="GI:912463"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04299.1"
 /db_xref="GI:912464"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04300.1"
 /db_xref="GI:912465"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04301.1"
 /db_xref="GI:912466"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04302.1"
 /db_xref="GI:912467"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04303.1"
 /db_xref="GI:912468"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04304.1"
 /db_xref="GI:912469"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04305.1"
 /db_xref="GI:912470"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04306.1"
 /db_xref="GI:912471"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04307.1"
 /db_xref="GI:912472"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04308.1"
 /db_xref="GI:912473"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04309.1"
 /db_xref="GI:912474"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04310.1"
 /db_xref="GI:912475"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04311.1"
 /db_xref="GI:912476"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04312.1"
 /db_xref="GI:912477"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04313.1"
 /db_xref="GI:912478"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04314.1"
 /db_xref="GI:912479"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04315.1"
 /db_xref="GI:912480"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04316.1"
 /db_xref="GI:912481"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04317.1"
 /db_xref="GI:912482"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04318.1"
 /db_xref="GI:912483"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04319.1"
 /db_xref="GI:912484"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04320.1"
 /db_xref="GI:912485"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04321.1"
 /db_xref="GI:912486"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04322.1"
 /db_xref="GI:912487"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04323.1"
 /db_xref="GI:912488"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAGYSAKXKPYASGS
 EMTIAGELRNGDALQANSVGLADYDEVNPFVFEPPSPHVSADGRPITDAALSDG
 IRLRLQADWLVLEGAGWFTPLSAEYFAFWVRQEQPLVLLVGIKLGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAFGRRRHOEYLAITLRMLPRCWAASRTCRPERAPLGO
 YDISLQAO"
 7215..7215
 /gene="bioc"
 7215..7215
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=1
 /product="the product of bioc"
 /protein_id="BA04324.1"
 /db_xref="GI:912489"
 /translation="MSKRRFVTGDTPEYKGSVASSALLQANRAG

Best Local Similarity 100.0%; Pred. No. 0.81;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtcgctgacatcgtcttgaagc 24
 |||
 Db 98 GTTGTGATCTCTCTTTGAAGC 121

RESULT 17

AE004192

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

gene
 CDS

1497. .2549
 /gene="VC1112"
 1497. .2549

gene
 CDS

1497. .2549
 /gene="VC1112"
 1497. .2549

gene
 CDS

1497. .2549
 /gene="VC1112"
 1497. .2549

gene
 CDS

1497. .2549
 /gene="VC1112"
 1497. .2549

gene
 CDS

transcriptional regulators e.g. glycine cleavage system
 transcription activator (gcv operon activator) -
 Escherichia coli: g1417043|s|p1932064|c|CVA_ECOLI percent
 identity 31 in 300 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative transcription regulator"
 /protein_id="BAB34178.1"
 /db_xref="GI:13360214"
 /translation="MRGKTPKTELLVTEVVAHREHYTRAELALTQSAVERQVSAL
 EEFKTPLESHSKRRKIFLNDGKYGLGKETLNKLEKDTINTMTPQVQVIELAVN
 PFEETHMLPMLHEFTKLHPDIIVNIHSLANNQPLNREYAVIMRENFCAPMAEVEY
 LFEEILPVCSSGLLMSDOKLSVALLETTELPLDHOSTRTIGMEWFAISGVSSPLVN
 NMPREDLSMLIAVRSNLGVALLPREFALIQHDLDGDMVPCDVPITRGNRFINTWCE
 EKSDSPHQOFREMLAKSVYQEM"
 complement(3720..4322)
 /gene="ECs0756"
 complement(3720..4322)
 /gene="ECs0756"
 /note="probable cob(I)atamin adenosyltransferase, similar
 to cob(I)atamin adenosyltransferases (corrinoid
 adenosyltransferases) e.g. [Escherichia coli]
 g1115148|s|p130401|BTRK_ECOLI percent identity 67 in 200
 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative cob(I)atamin adenosyltransferase"
 /protein_id="BAB34179.1"
 /db_xref="GI:13360215"
 /translation="MEARISTEHRVROOKLEQVDTRVAALIEKKGLIVFTGNGK
 KSTAFGIVTRAVGHGTGTVGAQYITKGOMNGEYNLQPLGVEFHITGFWETOR
 QADIDAKKEVSESKRLADKRDYLDVLDLTYLAHYHLDTEEVIASLGNRAQGV
 IVTRGCHSQILKADYSEIRPVKHAFFNMGIDQPGIDM"
 complement(4332..5984)
 /gene="ECs0757"
 complement(4332..5984)
 /gene="ECs0757"
 /note="probable fumarate hydratase, similar to fumarate
 hydratases e.g. fumarate hydratase class I, aerobic
 (fumarate) - Escherichia coli
 g1120598|s|p109231|FUMA_ECOLI percent identity 68 in 545
 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative fumarate hydratase"
 /protein_id="BAB34180.1"
 /db_xref="GI:13360216"
 /translation="MSKPFITWDEPFLONKDCTEYTLISDQHITVELDGEVEKIAPE
 ALTLARQATVEASFELRSALQOVASILNDQOASSNDKVALQILRNKEVSKVLP
 NCODTGAATVASKGQDITWGNDADALSKGIYTFQENNLRFSSQAPLDMSEVYTO
 TNLPAQIDISAVAGDEYHFLCVNKGSGSANKALYQETKSLQPEKLTAFLEKMSL
 GTAACPPYHIAFVVGSLADOTLKIAKLASTKYDNLPTSGNCOGAPFDELEKLL
 EASQOIGAGQGGKFAHDIRYLRPRHGSGPIMALSSCSNRIKAKINKHGTLW
 EKLEHNNGOYIPASLREHNAQVQDLDNPLRDVMDLARLDVGRVSLSGPTVAR
 DIAHAKTKARLDSEEPPEYELKHIVYVAGPAPKPNMAGSGAPTTGGRMDDYIDTF
 OAAGGSLVMLSKNRSQOVVDACHKGGFNLSIGAAALQOEVYKSLRCLEYPELG
 MEAVMMVEYENLPALFIVDDKGNFFSQFEQHRASCAPAGH"
 complement(6092..7372)
 /gene="ECs0758"
 complement(6092..7372)
 /gene="ECs0758"
 /note="probable transport protein, similar to
 glutamate/aspartate transport proteins (proton glutamate
 symport proteins) e.g. [Bacillus stearothermophilus]
 g1121467|s|p124943|GLUT_BACST percent identity 38 in 416
 aa, also similar to C4-dicarboxylate transport"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative transport protein"

/protein_id="BAB34181.1"
 /db_xref="GI:13360217"
 /translation="MRKISLTWMLLALVGLMIGVLNNTASPEPAKLYAOEISIFT
 TIFRLIKIMLIPLVSTLVGIAKMGDAKALGRISKTLFPLICASLLSLAIGLITV
 NFMPCGTGIFNFAVGAETGVVAPEPTKLVFSLHAFPSITYDMAAHNEIIDOIVFST
 ELCCSLTAIGKSAIVHALDLSAHMLKLTGVMLFAPLVFAAISALIRAGLAVM
 VSAKIFMGERTFTMLLWLVLTGLAIVVGPCTIRLTRALSSPALAFTSSSEAF
 GTLEKLEQFQVSPKIASFVLPITGVSNLVGSMAVCSFAVFTAOACNTHLSIGEDITM
 LLILMLTSKMGACVPRAHNVIAATLNQENIPEAGLILMGVDPLDMGRSATVMSN
 AMGAAMVSRWEGHFEQGRKALRPNESNALP"
 complement(7533..7850)
 /gene="ECs0759"
 complement(7533..7850)
 /gene="ECs0759"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34182.1"

Query Match 100.0%; Score 25; DB 1; Length 297816;
 Best Local Similarity 100.0%; Pred. No. 0.48;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgtagctcgtgttgaagcg 25
 Db 92367 GTTCTGCAATCTGCTGTTGAGCG 92391

RESULT 16
 E00893 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

E00893 1121 bp DNA PAT 29-SEP-1997
 Genomic DNA encoding biotin Synthetase.
 E00893.1 GI:2169154
 JP 1986149091-A/1.
 Escherichia coli.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 1121)
 Hirano, Y., Kojima, T. and Kimura, H.
 DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
 PRODUCTION OF BIOTIN
 Patent: JP 1986149091-A 1 07-JUL-1986;
 NIPPON SODA CO LTD
 OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIRONO YOSHIIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 C12N1:5/00, C12N1/20, C12P13/18, C12N1/20, C12P1:19, C12P13/18, PC
 C12N1:19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Ns101;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT CDS 42..1079
 /product="biotin synthetase".
 FEATURES
 source location/Qualifiers
 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 289 a 325 g 211 t
 ORIGIN

Query Match 96.0%; Score 24; DB 6; Length 1121;

```

/db_xref="GI:12513758"
/translation="MRPESLIVLIHQSSRSSMWIDLESLVASSYQWIKNNPOLRE
OGTHNATIQOILEALADKIGTETRTVLINLETSLRKSELYVIHLNSSPLUGWARD
IDAENHROGDFHCNLNAVYPGGEISVDVPATLVFHELLAVFHNLNGERAKES
SRAEOKSPPLLEEARTVYGAFSEEVLSENKFHELDIOMFRITSPXDSALIHDDVT
VSLGGQVRLHPLL"
/complement(7668..8144)
/gene="yhbh"
/feature="orf"
/note="Residues 1 to 158 are 99.36 pct identical to residues 1 to 158 of 158 from Escherichia coli K-12 Strain MO165: B0773"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISDLRQDRLPHRHVFVNGMCYDCDNLSPLAMWDVPAGTK
SFVVTCYDPDATGSGMMHWVNLPATRVLPQGFSGGLVAMPDEVLQPTTDGCKG
YDGAPRPGETIRYLFTVHALDVVERIDVEGASGAMGVNFHFSLASASTFAWF.S"
/complement(8203..9492)
/gene="bioA"
/note="20993"
/complement(8203..9492)

```

TITLE	Complete nucleotide sequence of the prophage virulence factor Shiga toxin 1 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain derived from the Sakai outbreak
JOURNAL	Gene 258 (1-2), 127-139 (2000)
MEDLINE	20564182
REFERENCE	4 (sites)
AUTHORS	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Toke,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.
TITLE	Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL	DNA Res. 8 (1), 11-22 (2001)
MEDLINE	21156231
REFERENCE	5 (bases 1 to 297816)
AUTHORS	Ohnishi,T., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/), Fax: 81-6-6879-2047
COMMENT	genome project.
FEATURES	Location/Qualifiers
SOURCE	1. 297816

RESULT	15
LOCUS	AP002553
DEFINITION	AP002553 297816 bp DNA
ACCESSION	AP002553
VERSION	AP002553.1
KEYWORDS	gi:13360211
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RLMD 0509952)
ORGANISM	DNA.
REFERENCE	Escherichia coli O157:H7
AUTHORS	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
TITLE	1 (sites) Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL MEDLINE REFERENCE	2 (sites) Ohnishi, M., Murata, T., Nakayama, K., Kohara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T. Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL MEDLINE REFERENCE	3 (sites) Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.

TITLE
Direct Submission
Submitted (22-Oct-2000) Laboratory of Genetics, University of
Journal
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Location/Qualifiers

source

1.13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"

misc-feature

<1..7576
/note="O-island #36: Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K: includes one copy of the 13 bp direct repeat that
flanks the prophage"

gene

66..665
/gene="lomK"
/note="20981"

CDS

66..665
/function="putative membrane; Other or unknown (Phage or
prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
g117532789|gb|AAE63231.1|AF151091_2 (AF151091) lom
[prophage P-Eiba]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"

/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKCAAILLSAICLAVSGVPAMASEHSTLSAGYLHASTDARG
SDDLNGIIVKRYEFTDTLGLTFSYANMEDQKTHSDTRHEDYVRRRMSYANG
PSRVNEMFSAYAMAGVAYSRVTSFGSDYPRDNRKRTDVLGSDARYSNTSLAM
GAGVQNPTESVANAVAYEXSGSDMRTDGLVGVTKP"
719..2041
/gene="20982"

gene

719..2041
/gene="20982"

/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 361 to 645 of 645 from Genpept 118 :
g114585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
tail fiber protein [Bacteriophage 933w]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"

/protein_id="AAG55138.1"
/db_xref="GI:12513753"

/translation="MGMVAVOISGLKDGAKPIQNCITOLKARRNSTXVVTFAVE
NPPEAGRYSMDEVGYSTVLIVGPPPSHAGTISYEDSOPCTLNDPLGAMFEDDAR
PEALRRPEQVVEEARHAEAKKKNAGFAEISARNAGISASKAASANATPSMEDSE
SARAASASAKKSEASSSASAEAKKSESLQSTDTDELSTKTAESAGNAADA
TTSREKARESAESQSKESRRIAEADVNIPLTVGPGGKGGPAGPGGDDGE
RGTPGAGATGEGPGGDTGPAGPGKGGRGEGTGLTGAGPGGPGGDTGAAGA
GPGGKGTGAAGVATGPGGPGGEGTGLTGAGPGGPGGDTGAAGA
TGLTLPKDATGVQWFOHLQVRFQDGPQWQVKGIDDEVGSDTGRGCE"
2004..2312
/gene="20984"

CDS

2004..2312
/gene="20984"

/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
g114585437|gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933w]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"

gene

/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MRKVAIOAELENDMLILKRLMOSLGGCKHDCENGRSLTAOLR
LGPADLIESDENGILPEQDRLVITQVVILLADKKQIQCVVRRPLQILRADGTWNIGSMK"
2489..3469
/gene="20985"

CDS

2489..3469
/gene="20985"

/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
g116960367|gb|AAE63527.1 (AF170176) hypothetical protein
predicted by Glimmer (Salmonella typhimurium LT2)"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPITRTYTHNSVNIQSSPCQVVSFAGKEVELKVIDEKPILF
QWPEPNPERKIKDEVPITVNTKQHPYLDNTNARLIESDMIGIFVGDPSVQKTAFS
KLERDEENWMIYREDVDPSMYDRKLSIDYHDIICFORLRTEDKRDVYLNLKEELR
EISKAODSLISMVAKKRNHAWDEPFRNLALAKAGEFPRCTYTKNHGTSFGGCTYLD
MMKILGKIGTITAPDGTSMHVDNRDSDNINENSATIVRSNHPALBGLSPMSKVD
AHRYVGLCKGVKKRYNFTPLNRYNHFCDFIEFNHPNITMTSQTSSW"
3530..4522
/gene="20986"

gene

3530..4522
/gene="20986"

CDS

3530..4522
/gene="20986"

/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"

/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG55141.1"

/db_xref="GI:12513756"

/translation="MKTPISLOSFPNFSAPAGSAPRIARRAENATADVLDIGRIPL
SAADISNYVESITRAVHDRSRRLIDQHYDMLIGNVLVDLSQTFRQAVSGLINKE
VHIGCTKYRNEYLENESSVKLIDDTIOSLTCNLTLYDVGOEIPFICAGENDDEPY
VSFSVAPDDDSYEMPSWDEGLTHEI1HHVTGSSDSGSDSNEILGPTTEILARRVOELG
WSVPDFKGYAEFEREAHLRLNMLNALROAMHEENBRFAEFERLGTISDRYEASPDF
EYSAAVSNIGYGTIOOHDFGLAINDNLQDANQIOLYHCAPYLFYFGDDKRNQO"
5350..6231
/gene="20989"

gene

5350..6231
/gene="20989"

/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"

/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG55142.1"

/db_xref="GI:12513757"

/translation="MLSPYSVNLGCSMNSLIRNTSPDNRVLSGRDAVSDMGAY
KGNTRYRVALDANKECYTRSHSCPTNILLRLGMPKSGELSKRTIEVNLASPVSAAM
REGIYHSNRDPLPDVYAPPELPSVDYRLSVPGVGVGKGNNAVYEDADATVLLAM
FTTSSQNEBVSIVGCFNMGVYAGASAEITYNNDDITGIRMDKINGSLNLSLPAO
AEHAIYVDFDRLEOKGILFVDTETNVLVYDRAKNEFNPIDISSYVSDRSWSESOIQ
SYHGKODLISVLSKI"
6462..7160
/gene="20990"

gene

6462..7160
/gene="20990"

CDS

6462..7160
/gene="20990"

/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"

/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG55143.1"

NNTYEGDVIVSGRAVVFNDNTEFVNSRGTQGEAVVEAPATLISNYVGTAVNSREN
 AEGDVACQSLSDVDANTNGOVVLRDSEAINEGFNTAKPWADAVISNRFPAQNTQSV
 DNDIQRNLNDINYNRMWYNNRGYSKVAARAKK"
 complement(1507. .1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 complement(1582. .2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582. .2058)
 /gene="ybhB"
 /function="orf: unknown"
 /note="f158; 99 pct identical to ybhB_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDCKLPHRHVFNCGYDGNISPHLAMDVPAGTK
 SFVYCTDPDAPFGSGMMWVNVNLPADTRVLPQSGSCIVAMPDGVLDTRIDFGKTS
 YDGAAPPKGETHRYIFVHALDIERIDVDGASGAMVCFVNHSLASITAMFS"
 complement(2108. .2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rnas predicted site"
 complement(2117. .3406)
 /gene="b10A"
 /note="b0774"
 complement(2117. .3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDDLAFDQRIHWPYTSPLPYVVSAGECGLLISDGR
 LVDSMSMAAIIHGYNHQNLAMKSOIDAMSHMFGGIHAAIELCRIIVMTDP
 LCEVLADSGSVAEVMKALQYQAKGRCFLFRNAGDFTFGASVCDPDS
 MSLMKGYLPENLFAFAPQSRMGEMDERMVGARLMAHREIAVILEPIVOGAG
 GMRKTHBWLKRITKICDRGILLIADIEATGNGRGKLFACSEAEIAPICIGKLU
 TGGTWTLSAILITREVAETISNCEAGCFMGPPTMCPRLACAAANSLALLESDDQ
 QVADIEVQLREQLAPARDAEIVADVRLGAIAGVETTHPVNMAALQKFVEQGVWIRP
 FGKLIIMPYIILPQOLQRLIAVNAVODEFFFCO"
 complement(2193. .2221)
 /gene="b10A"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411. .3450
 /note="central position to b10B promoter: -20"
 /bound_moiety="B10B documented site"
 3411. .3450
 /note="central position to predicted promoter: -20;genetic
 evidence for the site"
 /bound_moiety="B10B predicted site"
 complement(3411. .3450)
 /note="central position to predicted promoter:50"
 /bound_moiety="B10B predicted site"
 complement(3411. .3450)
 /note="central position to b10A promoter:50"
 /bound_moiety="B10B documented site"
 3413. .3441
 /note="factor Sigma70: promoter b10B: documented +1 at
 808525"
 complement(3447. .3473)
 /note="factor Sigma70: promoter b10A: documented +1 at
 808515"
 3493. .4533
 /gene="b10B"
 /note="b0775"
 3493. .4533
 /gene="b10B"
 /EC_number="2.8.1.1"

/function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0346; 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="Biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPWTLSQVTELEFEPKPLDLLPEAQOVHROHPDPOVOST
 LLSTKTAGCPEDCKYCPQSSRYKTGLEARLMEVEYLESAKAKAAGSTRFCMGAM
 KNPEHMDPYLEQWQVQKAMGLEACQKTLGTSQOURLANGLDYNNINLTSPF
 VGNLITRTYQERLDLEKVRDAGIKVCSGIVGLGTVADRAGLLQDLNLPSPS
 VPINLVYKGTPIADNDVDAPDFIRITAVARIMPTSYRLSAGREONNEGTQAC
 FMAGNSFVGCXKLITTPNPEDDQLFRKLGINPOOTAVLADNQQOQLRQALMT
 PPTDXYNAAL"
 4530. .5684
 /gene="b10F"
 /note="b0776"
 4530. .5684
 /gene="b10F"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0384; 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSMOEKIINALDARRADALRRYPVAGAGRWLVADROYLNF
 /SNDYGLSHHPOLIRAWOQGEQFGISGSGSYSVHQALELEBLAGYSRA
 LLFISGPAANOAVIAMAKEDRIALDEILSHASLLEASLSPSQIRFPAHNDYHAR
 LLASCPQGOQWYVTEGVSFMSGDSAPLAEIOYQOHNGMVLMDVDAHGTGIVGQGR
 SCWLQKVRPELLVTFGKGFVSGAAYLCSSVADYLIQFARHLIYTSMPAPQAOL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best local Similarity 100.0%; Pred. No. 0.34;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttgctgacatcgtcgttgaagcg 25
 Db 3549 gttgctgacatcgtcgttgaagcg 3573
 RESULT 14
 AE005258
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 AE005258 13501 bp DNA BCT 21-MAR-2001
 Escherichia coli O157:H7 EDL933 genome, config 1 of 3, section 82
 of 155.
 AE005258 AE005174
 AE005258.1 GI:12513751
 Escherichia coli O157:H7 EDL933.
 Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia
 1 (bases 1 to 13501)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)
 21074935
 11206551
 2 (bases 1 to 13501)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.


```

REFERENCE
  1 (bases 1 to 5872)
AUTHORS
  Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE
  Biotechnological method of producing biotin
JOURNAL
  Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
  source
    1. 5872
    /organism="unknown"
BASE COUNT
  1318 a 1552 c 1695 g 1307 t
ORIGIN
  1318 a 1552 c 1695 g 1307 t

Query Match
  Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
  1 gtgcgagatcgcgtcttgaagcg 25
  |||||||||||||||||||||||||
  173 gtgcgagatcgcgtcttgaagcg 197

RESULT 12
  ARI01810 5872 bp DNA PAT 14-FEB-2001
DEFINITION
  Sequence 6 from patent US 6083712.
ACCESSION
  ARI01810
VERSION
  ARI01810.1 GI:12812608
KEYWORDS
  Unknown.
SOURCE
  Unknown.
ORGANISM
  Unclassified.
REFERENCE
  1 (bases 1 to 5872)
AUTHORS
  Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE
  Biotechnological method of producing biotin
JOURNAL
  Patent: US 6083712-A 5 04-JUL-2000;
FEATURES
  source
    1. 5872
    location/Qualifiers
BASE COUNT
  1318 a 1552 c 1695 g 1307 t
ORIGIN
  1318 a 1552 c 1695 g 1307 t

Query Match
  Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
  1 gtgcgagatcgcgtcttgaagcg 25
  |||||||||||||||||||||||||
  173 gtgcgagatcgcgtcttgaagcg 197

RESULT 13
  AE000180 11022 bp DNA ECT 01-DEC-2000
DEFINITION
  Escherichia coli K12 MG1655 section 70 of 400 of the complete
  genome.
ACCESSION
  AE000180 U00096
VERSION
  AE000180.1 GI:1786988
KEYWORDS
  Escherichia coli K12.
  Escherichia coli K12.
  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Escherichia.
SOURCE
  1 (bases 1 to 11022)
AUTHORS
  Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
  Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
  Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
  Mau, B. and Shao, Y.
TITLE
  The complete genome sequence of Escherichia coli K-12
JOURNAL
  Science 277 (5331), 1453-1474 (1997)
MEDLINE
  97426617
PUBMED
  9278503
REFERENCE
  2 (bases 1 to 11022)

```

```

AUTHORS
  Blattner, F.R.
TITLE
  Direct Submission
JOURNAL
  Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
  University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
  Email: ecoliegnetics.wisc.edu Phone: 608-262-2534 Fax:
  608-263-7459
REFERENCE
  3 (bases 1 to 11022)
AUTHORS
  Blattner, F.R.
TITLE
  Direct Submission
JOURNAL
  Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
  University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
  Email: ecoliegnetics.wisc.edu Phone: 608-262-2534 Fax:
  608-263-7459
4 (bases 1 to 11022)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@amber.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated: this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Kiley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
  source
    location/Qualifiers
    1. 11022
    /organism="Escherichia coli K12"
    /strain="K12"
    /sub_strain="MG1655"
    /db_xref="taxon:83333"
    <1..130
    /note="REP (repetitive extragenic palindromic) element;
    contains 4 REP sequences"
    complement(147..1430)
    /gene="ybhC"
    /note="b0772"
    complement(147..1430)
    /gene="ybhC"
    /function="orf; Not classified"
    /note="1427: 98 pct identical to fragment YBHC ECOLI
    SW:P46130 (300 aa) but contains 127 additional C-terminal
    residues"
    /codon_start=1
    /transl_table=11
    /product="putative peptidesterase"
    /protein_id="AAC73859.1"
    /db_xref="GI:1786989"
    /translation="MNTFSVSLALALAFGYLTACSSPPDPPSPDPTAGTSSRPI
    LSAKEANPDADQHYFASLTPGAANPSPRLPQNPVPPAPATGCTHTTTIQAAYD
    AAIIRKTRKROIAYMPPEYQCTYVYPAAPGTTIDTIGKEKPIDVKTGLSLDGSMSPA
    DMRHYVNRGKTMPEKPAWYWDSCQSRSDISGLCSAVFWSQNNGLQNLNTIENI
    LGDSVDAGNHPVALRTDGDQVQINNNVILGRONTFVYVNSGVONRLTNPRTLVAT

```

REFERENCE	1 (bases 1 to 5872)
AUTHORS	Birch, O. and Brass, J.
TITLE	Biotechnological method of producing biotin
JOURNAL	Patent: EP 0798384-A 6 01-OCT-1997;
FEATURES	LOCUS AG (CH)
SOURCE	Location/Qualifiers
	1..5872
	/organism="Escherichia coli"
	/strain="DSM498"
	/db_xref="taxon:562"
	/clone="PRO30A15-9"
	1141..1156
RBS	/standard_name="B10F RBS"
gene	1154..2308
	/gene="B10F"
	1154..2308
CDS	/gene="B10F"
	/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
	/EC_number="2.3.1.47"
	/codon_start=1
	/transl_table=11
	/evidence=experimental
	/product="KAPA SYNTHASE"
	/protein_id="CAB69594.1"
	/db_xref="GI:6741868"
	/translation="MSMOEKINAIIDARBAADALRRRYIPVAGAGRMVLADPEYLNF
	SSNDYLDLSHPQITIRMOQGADEFGTSGSGSGHSGSYVHOALEETLAEMLGYSRA
	LLFTSGFANQAVIAAMAKEDRIADRLSHSLLEPAASLSFSQLRANDHTHLAR
	LLASPCCQOMVVYTEGVFSMDGSDAPLAEIQQVTOHNGMLVDDHGTGVIGDGGK
	SCWIOKPELLVYTEFGCGVSGAALVCSSTVADYLQFAHLIYTSMPADAQAL
	RASIAVIRSDGDAIREKILAIITRRAGVODLPFLADSCSAIOLPLIVGNSRALDL
	AEKIROGCGWTAIRPPTVAGTARLRILLTAAHEMODIDRLLEVILHNG"
	3030..3045
RBS	/standard_name="B10D RBS"
	3043..3753
gene	/gene="B10D"
	3043..3753
CDS	/gene="B10D"
	/standard_name="DETHIOBIOTIN SYNTHASE"
	/EC_number="6.3.3.3"
	/codon_start=1
	/transl_table=11
	/evidence=experimental
	/product="DTB SYNTHASE"
	/protein_id="CAB69595.1"
	/db_xref="GI:6741869"
	/translation="MSKRYEYVTDTEGKGTVASCALLQAAKAGYRTAGYKPYVAGSG
	ETKPEQDNLWDALQORNSLIQDLYATVNPYTPAETPSHIISAOGRPESIVNSAC
	LRALREQADWLVGAGGWFPLSDPTTFADWVTOQLPVTIVGKRLGICINHAMITP
	OYIOAGLTLTAGWADVPTPRKRNHALEYMTTLTRMLPA...GEIPLAENPENATGTG
	YINLAFVASTLIGTFSRL"
BASE COUNT	1318 a 1552 c 1695 g 1307 t
ORIGIN	
Query Match	100.0%: Score 25; DB 6; Length 5872;
Best Local Similarity	100.0%: Pred. No. 0.32;
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY	1 gttgtggtatcgcgtgttgaagcg 25
Db	173 gttgcttgatctgcgtttgaagcg 197
RESULT 11	
LOCUS	AR101809 5872 bp DNA PAT 14-FEB-2001
DEFINITION	Sequence 1 from patent US 6083712.
ACCESSION	AR101809
VERSION	AR101809.1 GI:12812607
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.

```

RESULT      8
LOCUS       A38251          5872 bp    DNA
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
            Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
            Escherichia coli.
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     other publication PL 308301 950724
            other publication CA 2145400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 9500809 950913
            other publication FI 951547 950331
            other publication JP 8501694T 960227.
FEATURES
source      location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02328.1"
            /db_xref="GI:22948307"
            /translation="MSWQEKINAIADARRADALRRRYPYAOGAGRWLVADROYLNF
            SSNDYGLSHHPQIIRAMQOAGQFGIGSGSGHSGSVYHQALEEELAEWLGYSRA
            LLIPTSGAANOAVIAAMAKEDRIADRLSHASILEASLSPQIRRFANDVTHLAR
            LLSPPGQOQNVVEGVFSMDGDSAPLAFIIOVTOCHNCLMDADAGTGVIGEOGKG
            SCWLOKYPPELLAVTFPGKGVGSAVLCSTVADYILDFRHLIYSTMPPAQOAL
            RSLAVIRSEDEGARREKLAALLTRFAGVODLPTFLADSCSAIOPLIVGDSRAIOL
            AEKLRQOGCWVTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNC"
            3030..3045
            /standard_name="BIOD RBS"
            3043..3373
            /gene="BIOD"
            3043..3373
            /gene="BIOD"
            /standard_name="DETHIOBIOTIN SYNTHASE"
            /EC_number="6.3.3.3"
            /codon_start=1
            /transl_table=11
            /number=4
            /evidence=experimental
            /product="DTB SYNTHASE"
            /protein_id="CAA02330.1"
            /db_xref="GI:2294851"
            /translation="MSKRYFVGTGTGTVEGKTVASCALLQAAKAGYRTAGYPVAGS
            EKTPGIRNSDALOIRNSLQIDATVNPYFAFPTSPHIIISAORPIESLVSAG
            LRALEQOADMVLEVGAGGFTPLSPFTFADWTDQPLVLLVGVKLCICINHAMITJA
            OVIOHAGLITLAVGVANDVYPPGKRHAETMTTLIRMIAPVLGELIPWLAENPENAAATGK
            YINLAPVDASTLGFTSRL"

```

```

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1      gtgtcgtgacgcgcgtcgtgaagc 25
Db      173      GTTGCTGATCTGCTGTTGAAGCG 197

RESULT      9
LOCUS       A93674          5872 bp    DNA
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
            Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
            Escherichia coli.
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES
source      location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="BIOD RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="BIOD"
            117..1157
            /gene="BIOD"
            /codon_start=1
            /transl_table=11
            /evidence=experimental
            /product="BIOTIN SYNTHASE"
            /protein_id="CAB69590.1"
            /db_xref="GI:6741863"
            /translation="MAHRPMTLSQVTELFEPKPLIDILFEAQOYRHOFPDPOYVST
            LKSIKGAQPEDCKYCPQSSRYKGTGLAEERMEYEOVLESRRKKAAGSTRFCGAAW
            KNIHERMDPYLEOMWOGYKAMGLPACHTLGLTLESORLNAAGLDYINHLDTSPF
            VGNITITRYOERLDTLKVADAGIKVCSGIVIGETVKDRAGLLIQNLNLPSPES
            VFINLVKVGPIADNDVDADFIRIIVARIIMPISYRISAGREONNEOTQAMC
            EMAGANSIFVCGCKLTLPNPPEBKDLQIFRKLGINPQOTAVLAGDNQOORLEQALMT
            PDIDRYNAAAL"
            2284..2297
            /standard_name="BIOD RBS"
            2295..3050
            /gene="BIOD"
            2295..3050
            /gene="BIOD"
            /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
            /codon_start=1
            /transl_table=11
            /product="PROTEIN"
            /protein_id="CAB69591.1"

```

BASE COUNT 1363 a 1554 c 1631 g 1245 t
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 1; Length 5793;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtctgagatcgcctgttgaagc 25
 |||||||
 Db 2068 GTTCTGATCTGCTGTTGAAGCG 2092

RESULT 7
 A38246 5872 bp DNA PAT 05-MAR-1997
 LOCUS Sequence 1 from Patent WO9408023.
 DEFINITION A38246
 ACCESSION A38246.1 GI:2294844
 VERSION
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
 LONZA AG (CH)

COMMENT
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 Location/Qualifiers

FEATURES
 source 1. 5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PRO30A-15/9"
 1. 96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23. 28
 45. 49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105. 119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117. 1157
 /gene="BIOB"
 117. 1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHRPRTLISOVTELEKPLDLIFENAOYHROHPDROYST
 LLSRTGACPEDCYQSSRYKGLAEERLMEVEVLESRRKAGSTRFCGAAW
 KNPERBMPYLEQWOCVAKMGLACMTLGTSQAKRLANGLDYYNNLDTSPE
 YGNITPTVOERDILEKVRDAGIKVSGGIVGLGTVDRAGLLQJLANLPTPPS
 VPIWIVKVGSTPLADNDVDAPDFIRIIVARIIMETSVRLSAGREONNEQOAMC
 FMAGNSFFYCKLITTPNPEEDKDLQFRKLGNPQOTAVLAGDNDQOQLFCALMT
 PDTEYVNAAL"
 2284. 2297
 /standard_name="BIOC RBS"

promoter
 -35_signal
 -10_signal

RBS

gene
 CDS

RBS

gene
 CDS

2295. 3050
 /gene="BIOC"
 2295. 3050
 /gene="BIOC"
 /function="INVOLVED IN PIMETOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVVKQALIAAFGRAAAHYEHQADIOROSADALLAMPORRYT
 HVLADACGPGQMSRHWERRHADVTLADLSPHILVOARODADHVLADIEPLATA
 TFDLAWSNLAWQWGNLSTALRELYKVRPRKVAFTLVQSGLEPHQAWAVDERP
 HANRFLPDEIFOSLNGVYOHRIQPTITLMPDVALSAMSLKIGATHLHEGRDRL
 TRSOLQRIOLAWPQOGRVPLTVHPLVGLARE"
 3742. 3752
 /standard_name="BIOA RBS"
 3750. 5039
 /gene="BIOA"
 3750. 5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE;
 8-AMINO-7-OXONONANOATE AMINOTRANSF."
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDDLAFFDQRIIMHPTYSMTSPLPYVVSABGCELLISDGR
 LVDSQSSMAAHHGYNHQPLNAAMKSOIDAMSHVMEGITHAPALICRLVAMTQP
 LECVELADSGVAVEAVKMALOYWQAKGEARQFLEFRNGYHGDFFGAMSVCDPNS
 MHSIKGCLPENLEFAPAPQSRMGEDEPRMNGFARLMAHREIAVITEPIYQAG
 GMRVHPEWLKRIKTCDBREGLILALDELINTCGRGKLFACCHAEIAPDILIGRAL
 TGCMTYLSATLTREVAETISNGEAGCFMGPPTPMGRPLACCAANSIATLIESGDMQ
 QVADIEVOLREOLAPARDAENVAVRVLGAIQVETHPVNMALQFFVEQCVIRP
 FGKLIYMPYIILPQOLQRLTAVNRAVODEFFFCO"
 5088. 5100
 /standard_name="ORF1 RBS"
 5098. 5574
 /gene="ORF1"
 5098. 5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGDKLPHRHVFNMGWYDGDNISPHLAWDVEAGTR
 STVVTCYDPDAPTCGSMHVVVNLPAPTRVLPQSGGLAMPDGVLOTPTDCKTG
 YDGAAPKGEHRTYLVHALDIEKIDVDEGASGAMGFNNHFIASASTAMFS"
 5583. 5605
 5583. 5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtctgagatcgcctgttgaagc 25
 |||||||
 Db 173 GTTCTGATCTGCTGTTGAAGCG 197

HLVDAGCGPGMWRHMRERHQAQVATLDSPEMLVOAROKDAADHYLAGDIESLPLATA
TFEDLAWSNLAIVOMCGNLSLALRELRYRVRKSGVAFPTLVQSGSLPELHOMAOVADERP
HANPELDPDEIEOSLNGVHOHITOPITLMDDDALSMBRLKIGATHLHEGNDPRL
TRSOLORLQALMPQOQGRPLTYHLFGVIARE"
4883..>5526
/gene="b10d"
/db_xref="GI:145426"
/gene="b10d"
/note="DTB synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase B10d"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRFVYGTDEYKTVASCALLQAAKAGRTAGYKPYAS
GSEKTEGLRNSDALALORNSILODYATVNPFTFAEPTSPHIIISAOEGRPISSVWS
SGTRLEQOADMVLEVCAGMFTPLSTFPAQVYTOEQLPVLIVGVKIGCTINHAHL
TAAQIQHAGLTLAGVAVANDVTPGKRHAETITTLTMTIPAPLIGEIPWLAENPE"
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN

Query Match 100.0%; Score 25; DB 1; Length 5526;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctgacatcctctgttgaagc 25
|||||
Db 2019 gttgctgacatcctgtttgaagc 2043

RESULT 6
ECOBIO 5793 bp DNA BCT 28-FEB-1994
LOCUS E.coli 7,8-diamino-pelargonic acid (b10a), biotin synthetase
DEFINITION (b10b), 7-keto-8-amino-pelargonic acid synthetase (b10f), bioc
protein, and dethiobiotin synthetase (b10d), complete cds.
ACCESSION J04423
VERSION J04423.1 GI:145422
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; b10a gene; b10b gene;
bioc gene; b10d gene; b10f gene; biotin synthetase; dethiobiotin
synthetase.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5793)
Otsuka,A.J., Buoncrisitanl,M.R., Howard,P.K., Flamm,J. and
Johnson,O.
The Escherichia coli biotin biosynthetic enzyme sequences
predicted
J. Biol. Chem. 263, 19577-19585 (1988)
J04423
J04423.1
Draft entry and computer-readable sequence [1] kindly submitted by
A.Otsuka, 09-NOV-1988.
FEATURES
source Location/Qualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=11
/protein_id="AAA23513.1"
/db_xref="GI:455168"
/translation="MKLISNDLRDQKLPRIHWPNGMGYDCDNI SPLHAMDVPACTK
SYVAATCDPADPAPGSGMHWVNVNLPADTRVLPOQFGSGLVAMPDGLQTRDPFGCTG
FSAAPPKGCTHRYIFTVHVALDIERIDVDEGASAAWGVFVHFLSASITMFS"
complement(633..1925)
/gene="b10a"
complement(633..1925)
CDS

/gene="b10a"
/codon_start=1
/transl_table=11
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AAA23514.1"
/db_xref="GI:457106"
/translation="MTTDLAFDORHIIWHPYTSMTSPLPVYPVVASCECELILSDGR
LVDSMSMAALIHGVNHPOLNAAMKSOJDAMSHVWFGITTHAPATELCRKIVAMSGN
ALECVFLASGSVAIVEVAMKALQYQAKGEARORFLEFRNGYHGDITGANSVCDPN
SMHSIMKGLPEPNLFAIPAPQSRMDGEWERDVGARLMAAHREHIAVILIEPYGA
GGMRMYHPMLKRIKICDREGILIDELATGFGRTGLFACHEHETADPILCLGKA
LVGCTMTLSATLITREVAETISNGAGCFMHPPTFMGNDPLACAANASTALIEGDWQ
OOVADIEVQLREOLAPARDAEMVADVIRVIGAGIVETTHPPNMAALQKFEVQGWIR
PGKILIMPPIIIIPQOQRLTAAVNNVADDETFEQ"
2012..3052
/gene="b10b"
2012..3052
/gene="b10b"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAA23515.1"
/db_xref="GI:145425"
/translation="MAHRRMTLSQVTELEFKPLDLLFEAQOYHRQHEDPRQOVST
LISIKTGACPEDCKYCPQTSRYKTGLEERLEVEVOVLESARAKAKAAGSTFFCGAAM
KNPHERDMPLYLEOMVOGVKAMGLEACMTLGTLSSEAOARLANAGIDYNNHLDPSP
YGNIIITRTYQERLDTLEKVRDAGIKVSGGIVGIGETVKRAGILLQALPLTPPS
VPINMLVKKYKGRPLADNDVDPAFDFIRITAVARIMPTSYRSLSGRCONBOQVAMC
PMAGANSTFYGCKLITTPREBDKDLQLEFKLGLNPQOTAVLAGDNEQORLEQALMT
PDTDEYVNAAL"
3049..4203
/gene="b10f"
3049..4203
/gene="b10f"
/codon_start=1
/transl_table=11
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AAA23516.1"
/db_xref="GI:145426"
/translation="MSMOEKINAALDARRADALRRRYPAQAGRMLVADPROLYNF
SSNDVLGSHNQITIRAMQOGEBOFTSGSGHVSIGSYVHQALEEELADWLYSRA
LLEISGFANQAVIAAMAKEDRIADRLSHASLLEASLSPSLRRRANHVDYHLAR
LLASPCQGOQWVTEGVFSMDGSDAPLAEIQVTOQHNGMLVNDABRTGVIGCGRG
SCMLQKVRPELLVTFFGKGVGSAVLCSTVADYLIQFRAHLIYTSMPDPAOAL
RASLAVIRDEGDARREKLALITFRFAGVODLPPTLADSCASLQIOPICVGNSSALQ
AKRLQOQGWVAIRPPVPACTARLRLITRAHMODIDRLLEYHNG"
4190..4945
/gene="b10c"
4190..4945
/gene="b10c"
/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVAKQAAIAAFGRAAAHVEQHADLQROSDALLMLPORKYT
HVLIDGCGGWSMRHMRERHQAQVATLDSPEMLVOAROKDAADHYLAGDIESLPLATA
TFDLAWSNLAIVOMCGNLSLALRELRYRVRKSGVAFPTLVQSGSLPERHQAQVADERP
HANRFLPDEIEOSLNGVHOHITOPITLMDDDALSMBRLKIGATHLHEGNDPRL
TRSOLORLQALMPQOQGRPLTYHLFGVIARE"
4938..5597
/gene="b10d"
4938..5597
/gene="b10d"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRFVYGTDEYKTVASCALLQAAKAGRTAGYKPYASGS
EKTPEGLNSDALALORNSILODYATVNPFTFAEPTSPHIIISAOEGRPISSVWS
LRALEQOADMVLEVCAGMFTPLSTFPAQVYTOEQLPVLIVGVKIGCTINHAHLTA
QVIOHAGLTLAGVAVANDVTPGKRHAETITTLTMTIPRCKWERSPLQIKMKOPES
T"

FEATURES	source
ORGANISM	Escherichia coli
REFERENCE	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
AUTHORS	1 (bases 1 to 1084)
JOURNAL	Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES	Location/Qualifiers
source	1..1084
gene	/organism="Escherichia coli"
CDS	/db_xref="taxon:562"
	24..1064
	/gene="BIOB"
	24..1064
	/gene="BIOB"
	/codon_start=1
	/transl_table=1
	/protein_id="CAA00965.1"
	/db_xref="GI:490219"
	/db_xref="SWISS-PROT:P12996"
	/translation="MAHRPHRTLSOVLELEFKPLDLLPEAOVHHROHPROVQVSNLSIKTAGCEPDCKYCCQSSRYKTKGLAEERLMEVEQVLESARKKAAKSTPCGAMAKNPERHMPYLEQWGVQKAWGLKACMTLGLTSSQARLANAGLDVYNHLDPSPERYGNITRTYKQERLDLLEKVDAGIKVCSGIVELGEGVRKADLLQLANLPPPEVPIMLYKVGKTPLANDDVAADFIRIYARIMPTSYRLSAGREONKEQVQAMFMAACSLFYQCKLLTTPNPEDKDLQFLRLGLNPGQTAVLADNDEQQQLLEQALMTPTDEYINAAAL"
BASE COUNT	271 a 286 c 318 g 209 t
ORIGIN	
Query Match	100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity	100.0%; Pred. NO. 0.27;
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 gtgtcgtgactcgtcttgaagc 25
Db	80 GTTGCTGCATCTCCTTTGAAGC 104
RESULT	5
AF250776	
LOCUS	AF250776 5526 bp DNA BCT 31-JAN-2001
DEFINITION	Uncultured bacterium pCosHE2. hypochlorite 17.1 kDa protein in moc-bioA intergenic region, DAPA-aminotransferase bioA (bioA), biotin biosynthesis bioB (bioB), KAPA synthetase bioF (bioF), and biotin biosynthesis protein bioC (bioC) genes, complete cds; and dehydrobiotin synthetase bioD (bioD) gene, partial cds.
ACCESSION	AF250776
VERSION	AF250776.1 GI:12620124
KEYWORDS	
SOURCE	
ORGANISM	uncultured bacterium pCosHE2.
REFERENCE	uncultured bacterium pCosHE2
AUTHORS	Bacteria: environmental samples.
TITLE	1 (bases 1 to 5526)
JOURNAL	Entcheva,P., Liebel,W., Johann,A., Hartsch,T. and Streitz,W.R.
MEDLINE	Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
PUBMED	20575196
REFERENCE	11133432
AUTHORS	2 (bases 1 to 5526)
TITLE	Entcheva,P., Liebel,W. and Streitz,W.R.
JOURNAL	Direct Submission
FEATURES	Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
source	Location/Qualifiers
	1..5526
	/organism="uncultured bacterium pCosHE2"
	/db_xref="taxon:143797"
	/clone="pCosHE2"
	/note="unknown organism, cosmid clone derived from environmental consortium"
	complement(52..5526)

gene
CDS
/note="ORF1"
/codon_start=1
/transl_table=1
/product="hypothetical 17.1 kDa protein in modc-b10A
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNLDRLDGDKLPNRRHVENSGMGDGINSPILADDDVPAGK
SPVYVCPDPAKPTGSGGSHVVVNLAPDTRLPDGPFGSGVAMPDGLQRTDPEGRKAG
YDGAAPRPGCEHRYIFYHALDIERIDVDEGASGAMGFNVHPSLASASITFMS"
complement(587..1876)
/gene="b10A"
/complement(587..1876)
/gene="b10A"
/note="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=1
/product="DAPA-aminotransferase B10A"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFDQRHIMHPYTSMTSPLPYIPVYSACECLISDGRRL
LVDGSSWMAAIGHYNNHQLNAAMSQIDASHVMFGSITHAPAIELCRKLVAMPPOP
LKVQSLAAGSGAVAAEVAKPMALQYQAQSGARORPLFRNRYHEDTFCGASVCDPDSNS
MHSLSKRGILPENLFAPQSRMDGEMDRDWGFAFLMAAREHETIAVYIIEPVIQVAGAG
GMRWHPMLKRIKRICDRREGILLADFLADGAFPGKTLFCACHEAELIATLIGKRL
TGGYITLSATLTFTREVATTISNGEAGCCMHQPTPMKGNPLACAAASATLIESDNQO
QVADIEVPLRLDAPDAEMVADRVLGALGVVEYTHPVMMALQNFVEQGWIRP
FCKLYIYMPITILPQOLRLTAAVNRVQDETFEQ"
1963..3003
/gene="b10B"
1963..3003
/gene="b10B"
/codon_start=1
/transl_table=1
/product="biotin synthase B10B"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPRTLISOVTELEFKPLLDLFEAQYHROHFDPRQYQVST
KLSTTGACPEDKCYQPQSSRYKTGLEERLMEVEQVLESARKAKAAGSTRETCGAAM
KNIRDPDTEOMQVQKDLGLEACMTLGLTSESQAORLANAGIDVYNNHLDLTPS
YGNITTTTQYQERLDTLEKVRDAGIKVCSGGIVIGETVKRAGILLQLANLPNPRES
VPINNLVVYQGLPDLNDNDVADPFIIRITIAARTIMPTSVYRLSGRPRQNEQOAMQ
EPAGANSIFYCGKLLITTPNPERDKDLQFKRLGLNPQOTAVLAGDNEQQLTSEALMT
PDTDYNNAAAL"
3000..4154
/gene="b10F"
3000..4154
/gene="b10F"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=1
/product="KAPA synthetase B10F"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="WSQEKIINALDARRAADALRRRIYPVAGQGRMLVADDRQYLN
SNSNDLQSHHPQILRAMQQAQEQFGTSGSGVSHQALEEELAWIDLYSRA
LLFISGFANQAVIAAMAKEDRIADRLSHASILEASLSPQLRRFANDVYLLHRA
LIASCPGOGLVTEGVSGMDQSDAPLAEIQGVQOONHGMILVMDAHGTQVIGQGRGR
SNQIQKVPDELLVTEFGKGFQSGAVALCSTSVADYILQFARHLIYTSMPAPDAQAL
RASLAVITRDEEDARREKLVSILIAHFRAGYQDDIPTTLADSCSAIOTPLVGNSSALQD
AEKLQDQCQWTAIRPTVPAGTARLRLITLTAHEMODIDRLLEYLHGNG"
4141..4896
/gene="b10C"
4141..4896
/gene="b10C"
/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=1
/product="biotin biosynthesis protein B10C"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="WATYNKCAIAAFGRMAALIEQHADLQROSDVLIAMLPERKITT

71	18.2	72.8	232744	3	AE003705	Drosophila
72	18.2	72.8	237119	2	AC017740	Drosophila
73	17.8	71.2	1571	8	SMCYPEG8	X71658 S. melongena
74	17.8	71.2	2250	3	DME277140	AJ277140 Drosophila
75	17.8	71.2	3569	3	CEU20168	U20168 Caenorhabditis
76	17.8	71.2	4166	1	MHU21963	U21963 Mycoplasma
77	17.8	71.2	4370	1	MHU21962	U21962 Mycoplasma
78	17.8	71.2	5358	1	AB030577	AB030577 Drosophila
79	17.8	71.2	7196	1	MHLMF	X81475 M. hominis 1
80	17.8	71.2	8580	1	AF083061	AF083061 Pseudomonas
81	17.8	71.2	20958	3	CEU04C10	Z69885 Caenorhabditis
82	17.8	71.2	29031	3	CEU05B10	Z74476 Caenorhabditis
83	17.8	71.2	32270	2	AC014470	Y12334 Takifugu ru
84	17.8	71.2	34602	2	CEFA5D3	Z78063 Caenorhabditis
85	17.8	71.2	37434	2	AC017697	AC017697 Drosophila
86	17.8	71.2	58378	2	AC016799	AC016799 Homo sapi
87	17.8	71.2	62868	2	AC007124	AC007124 Human gen
88	17.8	71.2	89263	9	AC006192	AC006192 Human DNA
89	17.8	71.2	89672	9	HS0822J19	AL031684 Human DNA
90	17.8	71.2	134784	2	AC061992	AC061992 Homo sapi
91	17.8	71.2	143712	9	HS917N8	AL031684 Human DNA
92	17.8	71.2	151611	9	AC009462	AC009462 Drosophila
93	17.8	71.2	154471	9	AC022523	AC022523 Homo sapi
94	17.8	71.2	155840	2	AC006495	AC006495 Drosophila
95	17.8	71.2	159813	2	AC010706	AC010706 Drosophila
96	17.8	71.2	167866	2	AC022994	AC022994 Homo sapi
97	17.8	71.2	176134	2	AC016182	AC016182 Homo sapi
98	17.8	71.2	176616	2	AC007449	AC007449 Homo sapi
99	17.8	71.2	181047	2	AC016765	AC016765 Homo sapi
100	17.8	71.2	192138	2	AC010997	AC010997 Homo sapi

ALIGNMENTS

RESULT 1	ECOBIOB	128 bp	DNA	BCF	03-JAN-1995
LOCUS	ECOBIOB	128 bp	DNA		
DEFINITION	Escherichia coli biotin (biob) gene, early terminator region.				
ACCESSION	M27731.1	GI:341755			
VERSION	M27731.1	GI:341755			
KEYWORDS	biob gene; biotin.				
SOURCE	Escherichia coli (strain K-12) DNA.				
ORGANISM	Escherichia coli				
REFERENCE	1 (bases 1 to 128)				
AUTHORS	Nath, S.K.				
TITLE	Attenuation of transcription of biotin genes in Escherichia coli				
JOURNAL	Can. J. Microbiol. 34 (12), 1288-1296 (1988)				
MEDLINE	89167942				
FEATURES	Location/Qualifiers				
source	1..128				
gene	/organism="Escherichia coli"				
-10_signal	/db_xref="taxon:562"				
mrna	4..128				
terminator	/gene="biob"				
	20..>128				
	/gene="biob"				
	103..121				
	/gene="biob"				
	/note="early terminator"				
BASE COUNT	31 a 28 c 34 g 35 t				
ORIGIN					

Query Match 100.0% Score 25; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtgttgaagcg 25
 Db 18 GTTGCTGATCTGCTGTTGAAGCG 42

RESULT 2
 LOCUS AR029499 1041 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 7 from patent US 5859335.
 ACCESSION AR029499
 VERSION AR029499.1 GI:5941472

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1041)
 AUTHORS Paton, D. Andrew.
 TITLE Enhanced biotin biosynthesis in plant tissue
 JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
 FEATURES location/Qualifiers
 source 1..1041

BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0% Score 25; DB 6; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtgttgaagcg 25
 Db 57 GTTGCTGATCTGCTGTTGAAGCG 81

RESULT 3
 LOCUS AR034916 1041 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 7 from patent US 5869719.
 ACCESSION AR034916
 VERSION AR034916.1 GI:5950521

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1041)
 AUTHORS Paton, D.A.
 TITLE Transgenic plants having increased biotin content
 JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
 FEATURES location/Qualifiers
 source 1..1041

BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0% Score 25; DB 6; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtgttgaagcg 25
 Db 57 GTTGCTGATCTGCTGTTGAAGCG 81

RESULT 4
 LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
 DEFINITION Biob gene of E.coli with primers.
 ACCESSION A11530
 VERSION A11530.1 GI:490218

KEYWORDS
 SOURCE Escherichia coli.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:55:14 ; Search time 1755.64 Seconds
(without alignments) 234.917 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25

Sequence: 1 gttgctgactctgctgttgaagcg 25

Scoring table: IDENTITY_NJC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824859755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pt:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_ov:*

21: em_pat:*

22: em_ph:*

23: em_pl:*

24: em_ro:*

25: em_sts:*

26: em_sy:*

27: em_un:*

28: em_vl:*

29: em_vl:*

30: em_htgo_hum:*

31: em_htgo_inv:*

32: em_htgo_rtd:*

33: em_htg_hum:*

34: em_htg_inv:*

35: em_htg_rtd:*

36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	25	100.0	128	1	ECOBIOB	M27731 Escherichia
2	25	100.0	1041	6	AR029499	AR029499 Sequence
3	25	100.0	1041	6	AR034916	AR034916 Sequence
4	25	100.0	1084	6	AI1530	AI1530 B108 gene o
5	25	100.0	5526	1	AF250776	AF250776 Unculture
6	25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
7	25	100.0	5872	6	A38246	A38246 Sequence 1
8	25	100.0	5872	6	A38251	A38251 Sequence 6
9	25	100.0	5872	6	A93674	A93674 Sequence 1
10	25	100.0	5872	6	A93679	A93679 Sequence 6
11	25	100.0	5872	6	AR101809	AR101809 Sequence
12	25	100.0	5872	6	AR101810	AR101810 Sequence
13	25	100.0	11022	1	AE000180	AE000180 Escherich
14	25	100.0	13501	1	AE005258	AE005258 Escherich
15	25	100.0	297816	1	AP002553	AP002553 Escherich
16	24	96.0	1121	6	E00893	E00893 Genomic DNA
17	20.8	83.2	12891	1	AE004192	AE004192 Vibrio ch
18	20.2	80.8	1041	1	EBH10TOP2	U38648 Erwinia her
19	20.2	80.8	7215	1	SMABIO	D17468 S. marcesce
20	19.8	79.2	8227	1	AF248314	AF248314 Unculture
21	19.8	79.2	185738	2	AC084015	AC084015 Homo sapi
22	19.8	79.2	227104	2	AC082387	AC082387 Homo sapi
23	19.8	79.2	302558	2	AC063928	AC063928 Homo sapi
24	19.8	79.2	303717	2	AC063930	AC063930 Homo sapi
25	19.2	76.8	22329	2	AC014347	AC014347 Drosophill
26	19.2	76.8	38000	9	AF025422	AF025422 Homo sapi
27	19.2	76.8	138938	2	AC010111	AC010111 Drosophill
28	19.2	76.8	176928	3	AC084239	AC084239 Homo sapi
29	19.2	76.8	181636	3	AC008224	AC008224 Drosophill
30	19.2	76.8	221420	2	AC011822	AC011822 Homo sapi
31	19.2	76.8	300050	9	HSMX1A	AL442166 Homo sapi
32	19.2	76.8	340000	9	HS21C084	AL163284 Homo sapi
33	18.8	75.2	550	10	AF158246	AF158246 Citicellul
34	18.8	75.2	106129	9	AC021088	AC021088 Homo sapi
35	18.8	75.2	165490	2	AC091895	AC091895 Homo sapi
36	18.8	75.2	175977	2	AC091721	AC091721 Pan trogl
37	18.8	75.2	229829	2	AC091717	AC091717 Pan trogl
38	18.6	74.4	4037	8	AFH249204	AJ249204 Arabidops
39	18.6	74.4	122877	9	AC034199	AC034199 Homo sapi
40	18.4	73.6	168091	2	AC016895	AC016895 Homo sapi
41	18.4	73.6	190000	9	AF213884S1	AF213884 Homo sapi
42	18.4	73.6	193980	9	AC021120	AC021120 Homo sapi
43	18.4	73.6	242015	2	AC092333	AC092333 Homo sapi
44	18.4	73.6	262608	2	AC092362	AC092362 Homo sapi
45	18.2	72.8	2018	3	AE002676	AE002676 Drosophill
46	18.2	72.8	2021	2	AC019735	AC019735 Drosophill
47	18.2	72.8	2094	5	XRMAHE	X57328 X.laevis mr
48	18.2	72.8	3429	10	AB029919	AB029919 Mus muscu
49	18.2	72.8	10405	1	AE003959	AE003959 Xylolla f
50	18.2	72.8	16644	1	AE006918	AE006918 Mycobacte
51	18.2	72.8	20760	1	MTCY21D4	Y80775 Mycobacte
52	18.2	72.8	38721	1	MSGY219	A_300013 Mycobacte
53	18.2	72.8	46300	8	NC8B4	A_353819 Neurospor
54	18.2	72.8	61162	2	AC087740	AC087740 Homo sapi
55	18.2	72.8	89795	8	NC3H10	AL513442 Neurospor
56	18.2	72.8	107868	9	AL445686	AL445686 Human DNA
57	18.2	72.8	118985	2	AC087152	AC087152 Mus muscu
58	18.2	72.8	129047	8	ATP23K16	AL078620 Arabidops
59	18.2	72.8	146349	2	AC007863	AC007863 Trypanoso
60	18.2	72.8	153099	9	CNS05TCE	AL355095 Human chr
61	18.2	72.8	155039	2	AC012548	AC012548 Homo sapi
62	18.2	72.8	161990	2	AL357313	AL357313 Homo sapi
63	18.2	72.8	166870	9	CNS05TEP	AL359228 Human chr
64	18.2	72.8	173328	2	AC019162	AC019162 Homo sapi
65	18.2	72.8	176475	2	AL355872	AL355872 Homo sapi
66	18.2	72.8	180934	2	AC092268	AC092268 Homo sapi
67	18.2	72.8	190539	2	AC025062	AC025062 Homo sapi
68	18.2	72.8	197019	9	AL157938	AL157938 Human DNA
69	18.2	72.8	198151	8	ATCHRIV91	AL161595 Arabidops
70	18.2	72.8	200799	2	AL445648	AL445648 Homo sapi

2728969-2733190: 25% of the driver population), and
 NCI CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550
 ; 25% of the driver population). Subtraction was
 performed as previously described [Bonaldi, Lennon &
 Soares (1996): Normalization and Subtraction: Two
 Approaches To Facilitate Gene Discovery. Genome Research
 6, 791-806.
 TAG_SEQ=None found"
 BASE COUNT 143 a 68 c 76 g 129 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 416;
 Best Local Similarity 83.3%; Pred. No. 6.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 ttgtcgcaagtcacagaattattt 25
 |||||
 Db 14 TTTTCTCAAGTCACATATTATT 37

Search completed: December 26, 2001, 12:11:11
 Job time: 6433 sec

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html) or from Reseach Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 144 row: 0 column: 14
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..335
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-14A014"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 102 a 55 c 52 g 126 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 335;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagattatt 24
||||| ||||||| ||| ||
Db 247 ATTGTCCAAAGTCACAAATAGTT 224

RESULT 44
A0019978 359 bp DNA GSS 09-JUN-1998
LOCUS CTF-HSP-2302N11.TF CTF-HSP Homo sapiens genomic clone 2302N11, DNA
DEFINITION sequence.
ACCESSION A0019978
VERSION A0019978.1 GI:3198714
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 359)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CTF-HSP-2302N11.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

Location/Qualifiers

1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2302N11"
/clone_lib="CTF-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 92 a 85 c 60 g 122 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 359;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagattatt 25
||||| ||||||| |||||||
Db 203 TTATCCAAAGTCACACTATTATT 226

RESULT 45
BF509677 416 bp mRNA EST 06-DEC-2000
LOCUS UI-H-B14-apb-d-06-0-01.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
DEFINITION IMAGE:3086651 3', mRNA sequence.
ACCESSION BF509677
VERSION BF509677.1 GI:11592975
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 416)
AUTHORS NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-rt@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www.bio.lnl.gov/bdtp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3086651"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8 is a subtracted library derived from NCI_CGAP-Sub5. The NCI_CGAP_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI_CGAP_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising a pool of clones from NCI_CGAP_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191, 254 of the driver population), a pool of clones from NCI_CGAP_Sub4 (IMAGE clone ids 2725592-2729326; 254 of the driver population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE ids

RESULT 39
 AZ456884/c 317 bp DNA GSS 04-OCT-2000
 LOCUS
 DEFINITION 1M0260B02F Mouse 10kb plasmid U0GCM library Mus musculus genomic
 clone U0GCM0260B02 F, DNA sequence.
 VERSION A2456884
 KEYWORDS A2456884.1 GI:10615009
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 317)
 Dunn,D., Aoyagi,A., Barber,M., Becorn,T., Duval,B., Hamll,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Kelly
 M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std. Error: 0.00
 Plate: 0260 row: B column: 02
 Seq primer: CTTGTAAACGACGCGCAGT
 Class: Plasmid ends
 High quality sequence stop: 317.
 Location/Qualifiers
 1. 317
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U0GCM0260B02"
 /clone_1lb="Mouse 10kb plasmid U0GCM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (g114732114[g0]A129072.1), a copy-number
 inducible derivative of plasmid p1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 69 a 64 c 76 g 108 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 317;
 Best Local Similarity 83.3%; Pred. No. 5.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtgcaggtacagaattatt 25
 ||||| ||||| ||||| |||||
 Db 64 TTGTCACAGTCACGAACTTTAT 41

RESULT 40
 A1495270/c
 LOCUS
 DEFINITION S026090.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl004-7673 5', mRNA sequence.
 VERSION A1495270
 KEYWORDS A1495270.1 GI:4396273
 SOURCE EST.
 ORGANISM soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 329)
 Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Coryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 T., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Riller,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 318.
 Location/Qualifiers
 1. 329
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl004-7673"
 /clone_1lb="Gm-cl004"
 /tissue_type="root"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
 XhoI; Root cDNA. The mRNA was isolated from entire roots
 of 8 day old 'Williams' seedlings which were propagated on
 paper towels with distilled water. Stratagene's cDNA
 Synthesis Kit (catalog #200401) was used to synthesize the
 cDNA. First strand synthesis was performed with 5-methyl
 dCTP, hence the ligated cDNA is hemimethylated.
 Stratagene's first-strand synthesis primer was used
 [GAGGAGCAGACAGAGACTGTCTCCAG(T)-18]. After
 second-strand synthesis, the cDNA ends were 'polished'
 with clone Pfu DNA polymerase, ligated to EcoRI adaptors,
 and phosphorylated. The XhoI site within the first-strand
 synthesis primer was restricted by digestion with XhoI;
 all XhoI sites in the cDNA would be protected by their
 hemimethylated status. The cDNA constructs were
 size-fractionated with a 500bp cutoff, using GIBCOBRL Life
 Technologies' cDNA Size Fractionation column. The column
 eluent was then ligated into Stratagene's pBluescript II
 XR Predigested vector (pBluescript II SK(+)) that had been
 digested with EcoRI and XhoI, and phosphorylated. Both
 the white and blue colonies appear to contain recombinant
 plasmids with cDNA inserts. Blue colonies (9n-15) have been
 sequenced, and possess putative cDNA inserts. This library
 was constructed by Dr. Paul Keim & Virginia H. Coryell,
 Department of Biology, Box5640, Northern Arizona
 University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.

FEATURES

source

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 557)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.

TITLE
JOURNAL
MEDLINE
COMMENT

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

FEATURES
source

Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.hsc.washington.edu
Plate: 902 Row: F Column: 22
Seq primer: 17
Class: BAC ends
High quality sequence stop: 557.
Location/Qualifiers
1..557

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=902 Col=22 Row=F"
/clone_lib="RPCI-11 Human Male BAC library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT 148 a 109 c 144 g 152 t 4 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 557;
Best Local Similarity 90.5%; Pred. No. 5.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 tctcgcaagtcacgaattat 23
||||| ||| ||||| |||||
Db 410 ttctcacactcagcagattat 390

RESULT 35
BF404390 646 bp mRNA EST 28-NOV-2000
LOCUS
DEFINITION UT-R-CAL-bie-f-08-0-UI.s1 UT-R-CAL Rattus norvegicus cDNA clone
ACCESSION BF404390
VERSION BF404390.1 GI:11392365
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 646)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 9704447
MEDLINE
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscores@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized corpus-striatum library cDNA library preparation. M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 382-596, >B3\$SINE/B2
Seq primer: M13 forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..646
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAL-bie-f-08-0-UI"
/clone_lib="UI-R-CAL"
/lab_host="PH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAL
library is a subcloned library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
, midbrain, cerebral cortex, corpus striatum, testis, and
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at latest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-CAL
TAG_TISSUE=corpus-striatum
TAG_SEQ=CTAGC"
BASE COUNT 159 a 117 c 132 g 238 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 646;
Best Local Similarity 90.5%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 5 ttcgaagtcacgaattat 25
||||| ||||| ||||| |||||
Db 160 ttccagacacagcagattatt 180

RESULT 36
AA927716 163 bp mRNA EST 17-JUN-1998
LOCUS
DEFINITION OM72105.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1552737 3',
mRNA sequence.
ACCESSION AA927716
VERSION AA927716.1 GI:3076536
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 163)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA library preparation: M. Bento Soares, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.

FEATURES

SOURCE

Location/Qualifiers
1. .965
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="074115"
/sex="male"
/note="Genoscope sequence ID: COBG074CP08LPI-end : T7"

BASE COUNT 285 a 170 c 193 g 316 t 1 others
ORIGIN

Query Match 72.8% Score 18.2; DB 13; Length 965;
Best Local Similarity 87.0% Pred. NO. 3.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgcgcagtcacagaattatt 25
||||| ||||| ||||| |||||
Db 565 TGTCTACTAGTCACGATTATTT 543

RESULT 32
A0070825/c 289 bp DNA GSS 05-AUG-1998
LOCUS
DEFINITION HS.2256 A1 F04 MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2256 Col=7 Row=K, DNA sequence.

ACCESSION A0070825
VERSION A0070825.1 GI:3389984
KEYWORDS GSS.

SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS I (bases 1 to 289)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.

TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

JOURNAL

99380589
Contact: Mahairas GC, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2256 Row: K Column: 7
Class: BAC ends
High quality sequence stop: 289.

FEATURES

SOURCE

1. .289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2256 Col=7 Row=K"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 117 a 76 c 39 g 57 t
ORIGIN

Query Match 71.2% Score 17.8; DB 13; Length 289;
Best Local Similarity 90.5% Pred. NO. 4.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 tgcgaatcacagaattatt 25
||||| ||||| ||||| |||||
Db 40 TCGCAGTCACAGATTATAT 20

RESULT 33

LOCUS

B1049694 317 bp mRNA EST 15-JUN-2001
CM2-GN0288-020101-667-d05 GN0288 Homo sapiens cDNA, mRNA sequence.

ACCESSION

B1049694
B1049694.1 GI:14457224

VERSION

KEYWORDS

SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 317)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

COMMENT

20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&t2=CM2-GN0288-
020101-667-d05&t3=2001-01-02&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 317.

Location/Qualifiers

FEATURES

SOURCE

1. .317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0288"
/dev_stage="Adult"
/note="Organ: placenta-normal; Vector: puc18; site_1: Sma1
; site_2: Sma1; A mini-library was made by cloning
products derived from OPRESS PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT 94 a 71 c 46 g 106 t
ORIGIN

Query Match 71.2% Score 17.8; DB 11; Length 317;
Best Local Similarity 90.5% Pred. NO. 4.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtgcgaatcacagaatt 21
||||| ||||| ||||| |||||
Db 186 ATTGTCAAGTACAGATT 206

RESULT 34

LOCUS

A0754255 557 bp DNA GSS 27-JUL-1999
HS_5326_B2.C11.T7A RPT1-11 Human Male BAC Library Homo sapiens

DEFINITION

genomic clone Plate=902 Col=22 Row=F, DNA sequence.

ACCESSION

A0754255
A0754255.1 GI:5618418

VERSION

KEYWORDS

SOURCE

human.
Homo sapiens

JOURNAL COMMENT

Unpublished (1998)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAATACGACCTACATATACGC
 Class: BAC ends
 High quality sequence stop: 372.

FEATURES

source

1..632
 Location/Qualifiers
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nxb0095A23f"
 /clone_lib="CUGI Rice BAC Library"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBeloBAC11. Site_1: HindIII. Site_2: HindIII. Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arunumathan and Farle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
 BASE COUNT 217 a 118 c 109 g 187 t 1 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 632;
 Best Local Similarity 87.0%; Pred. No. 3.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 tctgcgaagtcacagaattattt 25
 |||||
 Db 450 TGTCAACAGTCATAATATTATT 472

RESULT 30
 LOCUS AO200982 666 bp DNA 20-APR-1999
 DEFINITION RPI11-61E22.TK RPI1-11 Homo sapiens genomic clone RPI1-11-61E22.
 DNA sequence.
 ACCESSION AO200982
 VERSION AO200982.1 GI:3613181
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 666)
 Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Sub,E., Wible,C., de Jong,P. and Venter,J.C.
 Use of human BAC end sequences for Sequence-Ready Map Building
 JOURNAL Other_GSS: RPI11-61E22.TJ
 COMMENT Contact: Mark Adams
 Department of Eukaryotic Genomics

The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org

Library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/Bac_end_search/Bac_end_search.html
 Class: BAC ends.

FEATURES

source

1..666
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:7523157"
 /db_xref="taxon:9606"
 /clone="RPI1-11-61E22"
 /clone_lib="RPI1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPI11 Human Male BAC Library"
 BASE COUNT 224 a 117 c 148 g 175 t 2 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 666;
 Best Local Similarity 87.0%; Pred. No. 3.5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 tctgcgaagtcacagaattattt 25
 |||||
 Db 312 TGTGCAAGTTCACAGATATTATT 334

RESULT 31
 LOCUS CINS04111/c
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone 074L15 of library G from Tetradon nigroviridis, genomic survey sequence.

ACCESSION AL270334
 VERSION AL270334.1 GI:7992248
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 965)
 Roest-Crolius,H., Jallou,O., Dasilva,C., Fitzames,C., Fisher,C.,
 Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and
 Weissenbach,J.

Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 JOURNAL Unpublished
 TITLE 2 (bases 1 to 965)
 Roest-Crolius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fitzames,C., Wincker,P., Brotlier,P., Queller,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 JOURNAL Unpublished
 TITLE 3 (bases 1 to 965)
 Roest-Crolius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fitzames,C., Wincker,P., Brotlier,P., Queller,F.,
 Saurin,W. and Weissenbach,J.

Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.

	Matches	20;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
Qy	2	tttgcgaagtacagaattatt	24							
Db	325	TTGTCCCAAGTCACACAATTACT	303							

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L with: Mervin D, Gonsky R, Doctor

RESULT	28		
BG779264			
LOCUS	BG779264	622 bp	mRNA
DEFINITION	602665946F1	NIH_MGC_60	Homo sapiens cDNA clone IMAGE:4805891 5',

ACCESSION	mRNA sequence.
VERSION	BC779264
KEYWORDS	BC779264.1 GI:14049581
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

FEATURES	Location/Qualifiers
Source	1. .622

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 632)	Emmertloideae; Oryzaeae; Oryza.	
	Wing, R. A. and Dean, R. A.	A BAC End Sequencing Framework to Sequence the Rice Genome

/note="Vector: pT73D (Pharmacia) with a modified polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAGTGGAGCGCGGCTTTTCTTTTCTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT 39 a 45 c 36 g 49 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 169;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattat 23
||||| ||||||| |||||
Db 40 ATTGTCTCAGCTCACAGATTAT 62

RESULT 25
LOCUS AA920661 379 bp mRNA EST 20-APR-1998
DEFINITION vY48c01.r1 Striatagene mouse lung 937302 Mus musculus cDNA clone
IMAGE:1298688 5' similar to WP:F43E2.7 CE10348 ;, mRNA sequence.
ACCESSION AA920661
VERSION AA920661.1 GI:3067440
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 379)
Merrill, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:679736

Putative full length read
vector to vector length is 398
Seq primer: -28ml3 rev1 ET from Amerham
High quality sequence stop: 370.
Location/Qualifiers

FEATURES

SOURCE

1. 379
/organism="Mus musculus"
/strain="C57Bl/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1298688"
/clone_lib="Striatagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (Kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-2AP

XR Vector: -5' adaptor sequence: 5' GAATTGGCAGCAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3' "

BASE COUNT 102 a 69 c 87 g 121 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 379;
Best Local Similarity 87.0%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattat 23
||||| ||||||| |||||
Db 303 ATAGTCGCAATGCACAGAAATTAT 325

RESULT 26
LOCUS R21156 379 bp mRNA EST 18-APR-1995
DEFINITION yH20a04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:130254 5' , mRNA sequence.
ACCESSION R21156
VERSION R21156.1 GI:775937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 379)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Insert Size: 987
High quality sequence stops: 337
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 987 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 337.
Location/Qualifiers

FEATURES

source

1. 379
/organism="Homo sapiens"
/db_xref="GDB:537603"
/db_xref="taxon:9606"
/clone="IMAGE:130254"
/clone_lib="Soares placenta Nb2HP"
/sex="female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker: Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
AACTGAGAGATTCGCGCGCGAGATTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 123 a 56 c 83 g 117 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 379;
Best Local Similarity 87.0%; Pred. No. 3.2e+02;

REFERENCE	Acanthomorphidae: Acanthopterygii: Percomorpha: Tetraodoniformes: Tetraodontidae: Tetraodon.
AUTHORS	1 (bases 1 to 972) Roest-Crollius,H., Jallou,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A., and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 972) Roest-Crollius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Queller,F., Saurin,W. and Weissenbach,J.
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
TITLE	unpublished
JOURNAL	3 (bases 1 to 972)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon. location/Qualifiers
FEATURES	1..972 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="183108" /clone_lib="G" /note="genoscope sequence ID : COX6183BE04SP1-end : PUC-Orl"
SOURCE	BASE COUNT 262 a 196 c 218 g 266 t 30 others ORIGIN
Query Match	74.4% Score 18.6; DB 13; Length 972;
Best Local Similarity	84.0%; Fred. No. 2.4e+02;
Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy	1 attgcgcgaatcacagataattc 25
Db	164 ATTTCAGCAATTCACAGATTATT 140
RESULT	23
BF966737	BF966737 1015 bp mRNA EST 23-JAN-2001
LOCUS	602286714r1 NIH_MGC_95 Homo sapiens CDNA clone MIMAGE:4375802 3',
DEFINITION	mRNA sequence.
ACCESSION	BF966737
VERSION	BF966737.1 GI:12333952
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1015) NIH-MGC http://mhc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cragbs-remail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN) DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LHAM10041 row: k column: 03 High quality sequence start: 25

FEATURES					
High quality sequence stop: 735.					
Location/Qualifiers					
1..1015					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="IMAGE:4375802"					
/_clone_lib="NIH_MGC_95"					
/_tissue_type="hippocampus"					
/_lab_host="DH10B"					
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag...); Oligo-dr primed using primer 5'-TTTTTTTCTTTTTCVN-3'', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carrinci, in preparation). Library constructed by M. Brownstein (NIH/NIGRL, National Institutes of Health). Note: this is a NIH-MGC Library."					
BASE COUNT					
273 a 218 c 236 g 288 t					
ORIGIN					
Query Match 74.4%; Score 18.6; DB 11; Length 1015;					
Best Local Similarity 84.0%; Pred. No. 2.4e+02;					
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;					
OY	1	attgtcgaatgcacagaatatatt	25		
Db	455	AATGGCGGCTAGTGACACAATTATAT	479		
RESULT 24					
w18068 169 bp mRNA EST 10-SEP-1996					
mbd5d03.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone					
IMAGE:336197 5' similar to PIR:A48080 A48080 basic helix-loop-helix					
protein TFEC - rat ; , mRNA sequence.					
w18068					
w18068.1 GI:1292539					
EST.					
house mouse.					
Mus musculus.					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
1 (bases 1 to 169)					
Marra,M., Hillier,L., Allen,M., Bowles.M., Dietrich,N., Dubueque,T.,					
Geisell,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,					
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,					
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and					
Waterston,R.					
The WashU-HMI Mouse EST Project					
Unpublished (1996)					
Contact: Marra M/Mouse EST Project					
WashU-HMI Mouse EST Project					
Washington University School of MedicineP					
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108					
Tel: 314 286 1800					
Fax: 314 286 1810					
Email: mousetest@watson.wustl.edu					
This clone is available royalty-free through LINL ; contact the					
IMGE Consortium (info@image.llnl.gov) for further information.					
MGJ:217597					
Possible reversed clone: similarity on wrong strand					
Seq primer: mob.REGA+FW					
High quality sequence stop: 142.					
Location/Qualifiers					
1..169					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
/_clone="IMAGE:336197"					
/_clone_lib="Soares mouse p3NMF19.5"					
/_dev_stage="19.5 dpc total fetus"					
/_lab_host="DH10B (ampicillin resistant)"					
FEATURES					
SOURCE					

```

/db_xref="taxon:7668"
/clone_lib="Plate=1001 COL=23 Row=G"
urchin, sperm genomic BAC library"
/Note="Organ: Sperm; Vector: BACE3.6; BAC Clones in E-Coli
DH10B"
BASE COUNT      193 a      78 c      125 g      144 t      212 others
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 752;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 attgtcgcaagtcacagaattatt 25
        ||||| ||||| ||||| ||||| |||||
Db      89 ACTGTCAAGTCTCAGATCATTT 65

RESULT 20
BE548526      849 bp      mRNA      EST      09-AUG-2000
LOCUS      601075921F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3461861 5',
DEFINITION      mRNA sequence.
ACCESSION      BE548526
VERSION      BE548526
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 849)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-f@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
http://image.llnl.gov
Plate: LLAM8458 row: j column: 06
High quality sequence stop: 568.
Location/Qualifiers
1. 849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3461861"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/Note="Organ: Sperm; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT      248 a      226 c      159 g      216 t
ORIGIN

Query Match      74.4%; Score 18.6; DB 10; Length 849;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 attgtcgcaagtcacagaattatt 25
        ||||| ||||| ||||| ||||| |||||
Db      557 ACTGTCAAGTCTCAGATCATTT 533

RESULT 21
A0740253/c

```

```

LOCUS      A0740253      856 bp      DNA      GSS      16-JUL-1999
DEFINITION      HS.5505.A2.F08.77A.RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1081 COL=16 Row=K, DNA sequence.
ACCESSION      A0740253
VERSION      A0740253.1 GI:5517775
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 856)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GC, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1081 row: K column: 16
Seq primer: 77
Class: BAC ends
High quality sequence stop: 856.
Location/Qualifiers
1. 856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1081 COL=16 Row=K"
/clone_lib="RPCI-11 Human Male BAC library"
/sex="male"
/Note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      282 a      142 c      176 g      254 t      2 others
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 856;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 attgtcgcaagtcacagaattatt 25
        ||||| ||||| ||||| ||||| |||||
Db      552 ATTGTCAAGTCTCAGATCATTT 528

RESULT 22
CNS0220U/c
LOCUS      CNS0220U      972 bp      DNA      GSS      15-MAY-2000
DEFINITION      Tetradon nigroviridis genome survey sequence PUC-ori end of clone
183108 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION      AL221439
VERSION      AL221439.1 GI:7880258
KEYWORDS      GSS; genome survey sequence.
SOURCE      Tetradon nigroviridis.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Neoteleostei.

```


and selected for ampicillin resistance."

BASE COUNT 210 a 168 c 154 g 145 t

ORIGIN

Query Match 74.4%: Score 18.6; DB 13; Length 677;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgcgaagtcacagaattattc 25
||||| ||| ||||| ||| |||

Db 200 ATTGTCACACATCACAGATTGTTT 224

RESULT 15
B1222905/c 702 bp mRNA EST 11-JUN-2001
DEFINITION 602941167F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5104145 5',
mRNA sequence.
ACCESSION B1222905
VERSION B1222905.1 GI:14676349
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 702)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rt@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11251 row: f column: 18
High quality sequence stop: 701.
Location/Qualifiers
1..702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5104145"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 200 a 140 c 152 g 210 t

ORIGIN

Query Match 74.4%: Score 18.6; DB 11; Length 702;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgcgaagtcacagaattattc 25
||||| ||| ||||| ||| |||

Db 621 AATGGCGCTAGTCACAGATTATAT 597

RESULT 16
A0020375 709 bp DNA GSS 09-JUN-1998
LOCUS A0020375/c
DEFINITION CIT-HSP-2301C14.TRB CIT-HSP Homo sapiens genomic clone 2301C14, DNA
Sequence.
ACCESSION A0020375
VERSION A0020375.1 GI:3199111

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 709)
Adams M.D., Rounsley S.D., Zhao S., Field C.E., Bass S., Linher K.,
Golden K., Berry K., Granger D., Suh E., Wible C., Shizuya H.,
Simon M., and Venter J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other GSSs: CIT-HSP-2301C14.TPB
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..709
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2301C14"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11; Site: 1; HindIII; Site: 2;
HindIII"

BASE COUNT 303 a 109 c 124 g 173 t

ORIGIN

Query Match 74.4%: Score 18.6; DB 13; Length 709;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgcgaagtcacagaattattc 25
||||| ||| ||||| ||| |||

Db 664 ATTGTTGCAGTACAGATTTCCT 640

RESULT 17
BE394042 751 bp mRNA EST 21-JUN-2000
LOCUS BE394042/c
DEFINITION 601312405F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3659233 5',
mRNA sequence.
ACCESSION BE394042
VERSION BE394042.1 GI:9339407
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 751)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rt@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=989 Col=14 Row=L"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      167 a      126 c      109 g      168 t      6 others
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 576;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 atgttcgaagtcacagaattatt 25
||||| ||| ||| ||| ||| |||
Db 217 ATTGCTCTCAAGCCACACATTAAT 193

RESULT 13
A2083812      596 bp      DNA      GSS      08-MAY-2000
LOCUS
DEFINITION
RPC1-23-22N4..TJ RPC1-23 Mus musculus genomic clone RPC1-23-22N4,
DNA sequence.
ACCESSION
A2083812
VERSION
A2083812.1 GI:7725545
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
Zhao, S., Niemman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinet
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C. M.
Mus BAC End Sequences from Library RPC1-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tifg.org
Clones are derived from the mouse BAC library RPC1-23. For BAC
library availability, please contact Pieter de Jong
(pieter@edj.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tifg.org/tfb/bac_ends/mouse/bac_end_intro.html
Plate: 22 row: N column: 4
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..596
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-23-22N4"
/clone_lib="RPC1-23"
/sex="female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

```

```

BASE COUNT      174 a      108 c      95 g      219 t
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 596;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 atgttcgaagtcacagaattatt 25
||||| ||| ||| ||| ||| |||
Db 564 ATTGCTCAAGTCAAGACTAAT 588

RESULT 14
A2650467      677 bp      DNA      GSS      14-DEC-2000
LOCUS
DEFINITION
1M0520J07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0520J07 R, DNA sequence.
ACCESSION
A2650467
VERSION
A2650467.1 GI:11784986
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 677)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D. Weis, R.
Mouse whole genome scaffolding with paired end reads from 10Kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0520 row: J column: 07
Seq primer: CACACAGAAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 677.
FEATURES
Location/Qualifiers
1..677
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0520J07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-gold, T1-resistant, F-"
/note="Vector: PMW42mV; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMW42 (g11473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells

```


RESULT	8
N58785	
LOCUS	N58785 454 bp mRNA EST 30-JAN-1997
DEFINITION	yv76c09.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:248656 3' similar to contains Alu repetitive element;contains element l1 repetitive element ; , mRNA sequence.
ACCESSION	N58785
VERSION	N58785.1 GI:1202675
KEYWORDS	EST,
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 454)
AUTHORS	Hillier,L., Lennon,G., Becker,M., Ronaldo,M.F., Clapwell,B., Chissole,S., Dietrich,N., Dubnue,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,K.B., Morris,M., Parsons,J., Prange,C., Rikkin,I., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thillery-Meg,J., Trevasakis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996) 9704478
TITLE	Contact: Wilson RK
JOURNAL	Washington University School of Medicine
MEDLINE	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL : contact the IMAGE Consortium (infoimage.lnl.gov) for further information. Insert Length: 916 Std Error: 0.00 Seq primer: mob.REGA+ET High quality sequence stop: 1. Location/Qualifiers 1..454 /organism="Homo sapiens" /db_xref="GDB:3797902" /db_xref="taxon:9606" /clone="IMAGE:248656" /_clone_1lb="Soares fetal liver spleen INFLS" /_sex="male" /_dev_stage="20 week-post conception fetus" /_lab_host="DH10B (ampicillin resistant)" /_note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) With a modified polylinker. Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'] AACGTGAGAATTAATAAGACTTTTCTTTTTTTTTC 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Ronaldo."
BASE COUNT	116 a 65 c 104 g 168 t 1 others
ORIGIN	
Query Match	74.4%; Score 18.6; DB 11; length 454; Best Local Similarity 84.0%; Pred.No.2.2e+02;
Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	1 attgtcgcaatgcacagaattattt 25
Db	193 ATTGTTGCACGTAACAGATTTCF 217
RESULT	9
Bf711742/c	
LOCUS	Bf711742 459 bp mRNA EST 02-JAN-2001
DEFINITION	M1-P-E6-acfg-g-06-1-UM.s1 M1-P-E6 Sus scrofa cDNA clone M1-P-E6-acfg-g-06-1-UM 3', mRNA sequence.
ACCESSION	Bf711742
VERSION	Bf711742.1 GI:12011217

KEYWORDS	EST.
SOURCE	Pig.
ORGANISM	Sus scrofa
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE	1 (bases 1 to 459)
JOURNAL	Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE	Normalization and subtraction: two approaches to facilitate gene discovery
COMMENT	Genome Res. 6 (9), 791-806 (1996) 97044477 Contact: Tuggle CK Molecular Genetics Laboratory, Department of Animal Science Iowa State University 201 Kilgus Hall, Ames, IA 50011-3150, USA Tel.: 5152944252 Fax: 5152942401 Email: ckungjle@iastate.edu
FEATURES	The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized fetus at gestational day 20 library cDNA library Preparation: R1 Woods, JA Green, RS Praher S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com) Seq primer: M13 Forward POLYA=Yes. Location/Qualifiers 1..459 /organism="Sus scrofa" /strain="crossbred" /db_xref="taxon:9823" /clone="MI-P-E6-acf-g-06-1-Um" /clone.lib="MI-P-E6" /lab_host="DH10B (Life Technologies)" /vector="pTRF3D-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E6 library is derived from fetus at gestational day 20. For a detailed description of the library from which this clone was derived, please visit our web site at http://pigest.genome.iastate.edu/ TAG_LIB=MI-P-E6 TAG_TISSUE=fetus at gestational day 20 TAG_SEQ=AGGAA"
BASE COUNT	157 a 91 c 81 g 130 t
ORIGIN	
Query Match	74.4%; Score 18.6; DB 11; Length 459;
Best Local Similarity	84.0%; Pied. No. 2.2e+02;
Matches 21: Conservative	0; Mismatches 4; Indels 0; Gaps 0
QY	1 attctgcgaagtcacagaattatc 25 Db 385 atttcccaattcacacagtacttt 361
RESULT 10	
AZ277120/c	556 bp DNA GSS 26-JUL-2000
LOCUS	RPCI-23-134L7.TV RPCI-23 Mus musculus genomic clone RPCI-23-134L7,
DEFINITION	DNA sequence.
ACCESSION	AZ277120
VERSION	AZ277120.1 GI:9496111
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus. 1 (bases 1 to 556)

Db 296 ATTGTTGCAAGTCAGAGAGTCTTT 320
 ||||| ||||| ||||| ||||| |||||
 RESULT 6
 AM781150
 LOCUS 388 bp mRNA EST 12-MAY-2000
 s189g06.y1 Gm-cl037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl037-1235 5' similar to TR:082150 082150 CHLOROPLAST FRSH
 PROTEASE ; mRNA sequence.
 AM781150
 VERSION AM781150.1 GI:7795753
 KEYWORDS EST.
 SOURCE soybean
 ORGANISM Glycine max
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
 Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Phaseoleae:
 Glycine.
 1 (bases 1 to 388)
 REFERENCE
 AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Stepice, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 TITLE Public Soybean EST Project
 JOURNAL
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 375.
 Location/Qualifiers
 1..388
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
 /clone_1lb="Gm-cl037"
 /tissue_type="fully expanded leaves of greenhouse grown
 plants"
 /dev_stage="2 week old"
 /lab_host="DH10B"
 /note="Vector: pSPORT1. Site 1: NotI. Site 2: SalI. This
 cDNA library was constructed from mRNA isolated from fully
 expanded leaves of greenhouse grown plants that were 2
 weeks old. The library was prepared using the Lite
 Technologies Superscript cDNA library construction kit.
 Complementary DNA was synthesized from mRNA using a
 poly(dt) sequence with a NotI restriction site. SalI
 linkers adapters were ligated to the blunt-ended cDNA
 fragments followed by NotI digestion. The cDNA fragments
 were directionally cloned into the NotI-SalI restriction
 site of the pSPORT1 vector. The ligated cDNA fragments
 were transformed into E.coli Electro-Max DH10B host
 cells. This library was constructed in the laboratory of
 Dr. Lila Vodkin by Anu Khanna at the University of
 Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"
 BASE COUNT 123 a 67 c 92 g 106 t
 ORIGIN

Oy 1 atgtgcgaagtcacagaattattt 25
 ||||| ||||| ||||| ||||| |||||
 Db 329 ATACACGCCAAGTCACAGATTATGT 353
 ||||| ||||| ||||| ||||| |||||
 RESULT 7
 T83852 444 bp mRNA EST 16-MAR-1995
 LOCUS y066b07.s1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone
 IMAGE:113173 3' similar to gb|U87934|HUMANAL043 Human carcinoma
 cell-derived Alu RNA transcript, (cRNA); gb:XS4150.nal
 IMMUNOGLOBULIN ALPHA FC RECEPTOR PRECURSOR (HUMAN); contains Alu
 repetitive element; contains LI repetitive element ; mRNA sequence.
 T83852
 VERSION T83852.1 GI:712140
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 1 (bases 1 to 444)
 REFERENCE
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Treviski, E., Waterston
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 TITLE The WashU-Merck EST Project
 JOURNAL
 COMMENT Contact: Willson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 939
 High quality sequence stops: 334 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 939 Std Error: 0.00
 Seq primer: T3
 High quality sequence stop: 334.
 Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="GDB:468790"
 /db_xref="taxon:9606"
 /clone="IMAGE:113173"
 /clone_1lb="Soares fetal liver spleen INFUS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen. Vector: pT733D (Pharmacia)
 with a modified polylinker. Site 1: Pac I. Site 2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dt) primer
 [5' AACGCAAGATTAATTAAGATCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT733 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 108 a 69 c 100 g 162 t
 ORIGIN

Query Match 74.4%: Score 18.6; DB 10; Length 388;
 Best Local Similarity 84.0%: Pred. No. 2.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 atgtgcgaagtcacagaattattt 25
 ||||| ||||| ||||| ||||| |||||
 Db 190 ATTGTTGCAAGTCAGAGATTCTT 214

Query Match 74.4%: Score 18.6; DB 11; Length 444;
 Best Local Similarity 84.0%: Pred. No. 2.2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-879-8325
Fax: 06-879-8326.

FEATURES
source
Location/Qualifiers
1..239
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="81D04"
/clone_1ib="mouse embryonal carcinoma cell line F9"

BASE COUNT 52 a 46 c 64 g 77 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 239;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtgcgaagtcacagaattatt 25
||||| ||| ||||| |||||
Db 30 ATTGTCCCAATTACAGACTATT 6

RESULT 4
BE800706 357 bp mRNA EST 20-SEP-2000
LOCUS
DEFINITION
BS997006.Y1 Gm-cl049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl049-1091 5 similar to FR:082150 082150 CHLOROPLAST FTSH
PROTEASE.; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE800706
EST.
soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 357)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna
J., Bolla, B., Maitra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyle, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers
Y., Peterson, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schuck
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterson, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
putative full length read
vector to vector length is 364
High quality sequence stop: 340.
Location/Qualifiers
1..357

FEATURES
source

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl049-1091"
/clone_1ib="Gm-cl049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+, Site_1: EcoRI; Site_2:

XhoI: The Clark NIL was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from whole seedlings of 3 week old greenhouse
grown plants. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (GibcoBRL). The library was constructed
in cooperation with Dr. Paul Keim's Laboratory at Northern
Arizona University."

BASE COUNT 169 a 53 c 47 g 88 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 357;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtgcgaagtcacagaattatt 25
||||| ||| ||||| |||||
Db 133 ATACAGCAAGTCACAGATTATG 157

RESULT 5
AV525093 382 bp mRNA EST 01-SEP-2000
LOCUS
DEFINITION
AV525093 Arabidopsis thaliana aboveground organs two to six-week
old Arabidopsis thaliana cDNA clone APD17c09 5', mRNA sequence.
AV525093
VERSION
KEYWORDS
SOURCE
ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 382)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
Location/Qualifiers
1..382

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APD17c09R"
/clone_1ib="Arabidopsis thaliana aboveground organs two to
six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 101 a 87 c 89 g 105 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 382;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtgcgaagtcacagaattatt 25

86 17.2 68.8 353 10 AA648449
 C 87 17.2 68.8 375 11 BF083176
 C 88 17.2 68.8 376 11 BF083167
 C 89 17.2 68.8 376 11 BF083177
 C 90 17.2 68.8 376 13 AF0882180
 C 91 17.2 68.8 382 11 N77748
 C 92 17.2 68.8 389 10 AA069844
 C 93 17.2 68.8 390 10 BE606001
 C 94 17.2 68.8 396 11 B1030143
 C 95 17.2 68.8 402 10 AV124671
 C 96 17.2 68.8 405 11 BF726110
 C 97 17.2 68.8 406 10 AA119160
 C 98 17.2 68.8 457 10 A1161386
 C 99 17.2 68.8 471 10 A1635259
 C 100 17.2 68.8 483 10 AW202805

ALIGNMENTS

RESULT 1
 BE540498/c 866 bp mRNA EST 09-AUG-2000
 LOCUS 601066628f1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:3452959 5',
 DEFINITION mRNA sequence.
 ACCESSION BE540498.1 GI:9769143
 VERSION BE540498
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Email: Robert Strausberg, Ph.D.
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.llnl.gov
 Plate: LAM8435 row: g column: 08

FEATURES
 source
 1.866
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3452959"
 /clone_lib="NIH_MGC_10"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
 Average insert size 1.5 kb. Library prepared by Life
 Technologies."

BASE COUNT 255 a 134 c 182 g 294 t 1 others

Query Match 76.8%; Score 19.2; DB 10; Length 866;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtgcaggtacagattatt 25
 Db 642 ttgtcccgagtcacacattatt 619

RESULT 2

AA0897324/c 547 bp DNA GSS 10-NOV-1999
 LOCUS HS_3120_B2-B04-T7C CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3120 Col=8 Row=D, DNA sequence.
 ACCESSION AA0897324
 VERSION AA0897324.1 GI:6353514
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 Contact: Mahairas GG, Wallace JC, Hood L
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.husc.washington.edu

Plate: 3120 row: D column: 8
 Seq Primer: T7
 Class: BAC ends
 High quality sequence stop: 547.

FEATURES
 source
 1.547
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3120 Col=8 Row=D"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in
 E-coli DH10B"

BASE COUNT 127 a 111 c 94 g 210 t 5 others

Query Match 75.2%; Score 18.8; DB 13; Length 547;
 Best Local Similarity 90.9%; Pred. No. 1.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 ttgtgcaggtacagattatt 24
 Db 354 ttgtcccgagtcacacattatt 333

RESULT 3
 D21670/c 239 bp mRNA EST 07-OCT-1996
 LOCUS MMS81D04 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
 DEFINITION clone 81D04, mRNA sequence.
 ACCESSION D21670
 VERSION D21670.1 GI:618798
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.
 TITLE A survey of genes expressed in undifferentiated mouse embryonal
 carcinoma F9 cells: characterization of low-abundance mRNAs
 J. Biochem. 116, 128-139 (1994)

JOURNAL MEDLINE
 COMMENT Contact: Kazunori Shimada

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:03 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-2
Perfect score: 25
Sequence: 1 attgtgcacgtcacagaattatt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

EST.*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_lin:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.2	76.8	866	10	BE540498 601066628
C 2	18.8	75.2	547	13	A0897324 HS_3120_B
C 3	18.6	74.4	239	11	D21670 MUSA1D04 mc
C 4	18.6	74.4	357	11	BE800706 sq97c06.y
C 5	18.6	74.4	382	10	AV525093 AV523093
C 6	18.6	74.4	388	10	AM781150 s189906.y
C 7	18.6	74.4	444	11	TS8852 yd66b07.s1
C 8	18.6	74.4	454	11	NS8785 yv76c09.s1
C 9	18.6	74.4	459	11	BF711742 MT-P-E6-e
C 10	18.6	74.4	556	13	AZ277120 RPCI-23-1
C 11	18.6	74.4	563	13	AZ207443 SP_0135-A
C 12	18.6	74.4	576	13	A0771996 HS_5413_H

13	18.6	74.4	596	13	AZ083812 RPCI-23-2
14	18.6	74.4	677	13	AZ650467
C 15	18.6	74.4	702	11	B1222905 602941167
C 16	18.6	74.4	709	13	A0020375 CIT-HSP-2
C 17	18.6	74.4	751	10	BE394042 601312405
C 18	18.6	74.4	751	10	BE658717 GM700007A
C 19	18.6	74.4	752	13	AZ183578 SP_1001_A
C 20	18.6	74.4	849	10	BE548526 601075921
C 21	18.6	74.4	856	13	A0740253 HS_5305_A
C 22	18.6	74.4	972	13	CNS02200 Tetradon
C 23	18.6	74.4	1015	12	BF966737 602286714
C 24	18.2	72.8	169	11	W18068 mb58403.r1
C 25	18.2	72.8	379	11	AA920661 vY48c01.r
C 26	18.2	72.8	379	11	R21156 yb20e04.r1
C 27	18.2	72.8	441	13	A0885836 HS_5527_B
C 28	18.2	72.8	622	11	BG779264 60265946
C 29	18.2	72.8	632	13	A0510232 nbxb0095A
C 30	18.2	72.8	666	13	A0200982 RPCI11-61
C 31	18.2	72.8	965	13	CNS041L1 Tetradon
C 32	17.8	71.2	289	13	A0070825 HS_2256_A
C 33	17.8	71.2	317	11	B1049694 CM2-GN028
C 34	17.8	71.2	557	13	A0754255 HS_5326_B
C 35	17.8	71.2	646	11	BF404390 UI-R-CA1
C 36	17.6	70.4	163	10	AA927716 om72f05.s
C 37	17.6	70.4	276	11	H60887 YF46F09.r1
C 38	17.6	70.4	300	11	C34741 C34741 Yuj1
C 39	17.6	70.4	317	13	AZ456884 1M0260B02
C 40	17.6	70.4	329	10	A1495270 sb02909.Y
C 41	17.6	70.4	335	11	H75480 yu61d12.s1
C 42	17.6	70.4	335	11	230954 APTS2670 Ve
C 43	17.6	70.4	335	13	AZ289215 RPCI-23-1
C 44	17.6	70.4	359	13	A0019978 CIT-HSP-2
C 45	17.6	70.4	416	11	BE509677 UI-H-B14
C 46	17.6	70.4	420	13	A0008647 CIT-HSP-2
C 47	17.6	70.4	432	11	N20061 yx28f07.s1
C 48	17.6	70.4	445	13	A0463730 HS_5051_B
C 49	17.6	70.4	450	10	AM936817 sg53e10.Y
C 50	17.6	70.4	463	10	A1275535 q163g05.x
C 51	17.6	70.4	480	10	AM934345 EST360188
C 52	17.6	70.4	482	11	N73875 z662a06.s1
C 53	17.6	70.4	483	13	A0973386 RPCI-23-3
C 54	17.6	70.4	494	11	BG656403 1b37a03.x
C 55	17.6	70.4	502	13	AZ234341 RPCI-23-6
C 56	17.6	70.4	520	13	A0669162 HS_5397_B
C 57	17.6	70.4	538	11	B1210793 EST528833
C 58	17.6	70.4	541	11	BE508978 UI-H-B14
C 59	17.6	70.4	543	13	A0778373 HS_3161_B
C 60	17.6	70.4	563	13	AZ249546 RPCI-23-7
C 61	17.6	70.4	567	13	AZ027939 RPCI-23-3
C 62	17.6	70.4	575	11	N52629 yv37g12.s1
C 63	17.6	70.4	577	13	AZ010791 RPCI-23-2
C 64	17.6	70.4	584	10	AM040981 EST83845
C 65	17.6	70.4	601	13	AZ394303 1M0157L18
C 66	17.6	70.4	604	10	AM550934 CM210009B
C 67	17.6	70.4	619	13	AZ256327 RPCI-23-1
C 68	17.6	70.4	622	13	BH125940 RPCI-24-3
C 69	17.6	70.4	634	10	AM965387 EST37505
C 70	17.6	70.4	643	13	A0306368 HS_2050_A
C 71	17.6	70.4	656	13	A0008644 CIT-HSP-2
C 72	17.6	70.4	659	10	AM038199 EST79856
C 73	17.6	70.4	683	10	AM998515 EST309015
C 74	17.6	70.4	694	13	A0332879 RPCI-11-3
C 75	17.6	70.4	694	13	AZ687232 ENTUX39TF
C 76	17.6	70.4	935	11	B1327624 602979116
C 77	17.6	70.4	944	10	A1529608 u180b10.Y
C 78	17.6	70.4	951	13	AL286564 Tetradon
C 79	17.6	70.4	1077	13	AL188397 Tetradon
C 80	17.4	69.6	495	10	BE756065 210103 MA
C 81	17.4	69.6	861	11	BF677689 602085574
C 82	17.4	69.6	887	13	AL212178 T3 end of
C 83	17.2	68.8	301	10	AA507304 nh49d02.s
C 84	17.2	68.8	348	11	R88883 yP97a02.r1
C 85	17.2	68.8	352	10	AA447045 zw86a10.r

MOLECULE TYPE: cdna to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: CDS
LOCATION: 144..2861
US-08-687-379-1

Query Match 61.6%; Score 15.4; DB 1; Length 2946;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacaga 18
|||||
Db 279 TTGTGCGCACTCACAGA 295

RESULT 44
US-08-687-379-3
Sequence 3, Application US/08687379
Patent No. 5756697
GENERAL INFORMATION:
APPLICANT: Hoefer, Thomas
APPLICANT: Ullsch, Andreas
APPLICANT: Bach, Alfred
APPLICANT: Sterrer, Sylvia
APPLICANT: Lemaire, Hans-Georg
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Keil & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna to mRNA
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: CDS
LOCATION: 144..2861
US-08-687-379-3

Query Match 61.6%; Score 15.4; DB 1; Length 2946;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacaga 18
|||||
Db 279 TTGTGCGCACTCACAGA 295

RESULT 45
US-08-254-573-1
Sequence 1, Application US/08254573
Patent No. 5610032
GENERAL INFORMATION:
APPLICANT: RAMBOU, Rajender
APPLICANT: ELIOTT, Candace
APPLICANT: NOTT, Stephen
TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,573
FILING DATE: 06-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,611
FILING DATE: 10-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/179 ALLE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3220 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 62..2782
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 62..115
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 116..2782
US-08-254-573-1

Query Match 61.6%; Score 15.4; DB 1; Length 3220;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacaga 18
|||||
Db 197 TTGTGCGCACTCACAGA 213

Search completed: December 26, 2001, 12:20:11
Job time: 6968 sec

SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/980,329C
FILING DATE: No. 5994118ember 28, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: Quest 4.1-154
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5994118e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2215
TYPE: Nucleotide
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE:
DESCRIPTION: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N/A
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN:
INDIVIDUAL ISOLATE: W1
DEVELOPMENTAL STAGE: N/A
HAPLOTYPE: N/A
TISSUE TYPE: N/A
CELL TYPE: bacterium
CELL LINE: N/A
ORGANELLE: N/A
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: SMO-21
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: phage abortive infection
LOCATION: N/A
IDENTIFICATION METHOD: sequencing
OTHER INFORMATION: DNA encoding phage resistance
PUBLICATION INFORMATION: N/A
US-08-980-329C-1

Query Match 61.6%; Score 15.4; DB 2; Length 2215;
Best Local Similarity 76.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 atgtgcgaagtcacagaattatt 25
||| ||| ||| ||| ||| ||| |||
Db 1925 ATTTCGAAGTTCATATTATTTT 1901

RESULT 42
US-08-904-284-6
Sequence 6, Application US/08904284
Patent No. 6133435
GENERAL INFORMATION:
APPLICANT: Fernandez, Donna E.
APPLICANT: Heck, Gregory R.
TITLE OF INVENTION: EXPRESSION OF AGL15 SEQUENCE IN
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison

STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/904,284
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296.94193
TELEPHONE: (608) 251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2437 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-904-284-6

Query Match 61.6%; Score 15.4; DB 3; Length 2437;
Best Local Similarity 76.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 atgtgcgaagtcacagaattatt 25
||| ||| ||| ||| ||| ||| |||
Db 2410 ATCTTACAGCAGACAGATTATT 2434

RESULT 43
US-08-687-379-1
Sequence 1, Application US/08687379
Patent No. 5756697
GENERAL INFORMATION:
APPLICANT: Hoeger, Thomas
APPLICANT: Ullsch, Andreas
APPLICANT: Bach, Alfred
APPLICANT: Stetler, Sylvia
APPLICANT: Lemaire, Hans-Georg
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kell & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-916-443A-15

Query Match 61.6%; Score 15.4; DB 3; Length 946;
Best local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagaattattt 25
||| ||||| ||||| ||
Db 699 ATGCTTCAAGTCGAGAAATTAGTT 723

RESULT 39
US-08-904-284-1
; Sequence 1, Application US/08904284
; Patent No. 6133435
; GENERAL INFORMATION:
; APPLICANT: Fernandez, Donna E.
; TITLE OF INVENTION: EXPRESSION OF AGL15 SEQUENCE IN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,284
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-904-284-1

Query Match 61.6%; Score 15.4; DB 3; Length 1070;
Best local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagaattattt 25
||| ||||| ||||| ||
Db 1038 ATCTTACAGGCACAGATTATT 1062

RESULT 40
US-08-916-443A-16
; Sequence 16, Application US/08916443A
; Patent No. 6001986
; GENERAL INFORMATION:
; APPLICANT: Yong Sig KIM

APPLICANT: Sun Chung PARK
APPLICANT: Soo Kyung OH
APPLICANT: Hosul LEE
APPLICANT: Jeong Woo CHO
APPLICANT: Chang H. CHUNG
TITLE OF INVENTION: Antiviral Proteins, Amarandin 1 and 2, from
TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,443A
FILING DATE: 22 AUG 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 1942/18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1116 nucleic acids
TYPE: nucleotides
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-916-443A-16

Query Match 61.6%; Score 15.4; DB 3; Length 1116;
Best local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagaattattt 25
||| ||||| ||||| ||
Db 858 ATGCTTCAAGTCGAGAAATTAGTT 882

RESULT 41
US-08-980-329C-1/C
; Sequence 1, Application US/08980329C
; Patent No. 5994118
; GENERAL INFORMATION:
; APPLICANT: Sylvain Molneau, Eric Emond,
; APPLICANT: Shirley A. Walker, Ebenezer R. Vedamuthu,
; APPLICANT: Jeffrey K. Kondo
; TITLE OF INVENTION: DNA Encoding Phage
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette 5.25 inch,
MEDIUM TYPE: 360 kb storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS (version 4)


```

: TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
: TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: 1601 Market Street, 36th Floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103-2398
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,420
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/374,843
: FILING DATE: 18-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary, Kathryn Ph.D.
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 567-2020
: TELEFAX: (215) 567-2991
: TELEX: 831-494
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 450 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-905-420-18

Query Match          62.4%  Score 15.6;  DB 2;  Length 450;
Best local Similarity 81.8%;  Pred. No. 84;
Matches 18;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY  2 ttgtcgcaagtcacagaattat 23
      |||||  |||  |||  |||
Db  260 TTGTGCAAGTCACATAGTTAT 281

RESULT 32
US-08-374-843B-9
: Sequence 9, Application US/08374843B
: Patent No. 5726016
: GENERAL INFORMATION:
: APPLICANT: Demuth, Donald R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
: TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: 1601 Market Street, 36th Street
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103-2398
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,420
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/374,843
: FILING DATE: 18-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary, Kathryn Ph.D.
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 567-2020
: TELEFAX: (215) 567-2991
: TELEX: 831-494
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 992 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 200..652
: US-08-374-843B-9
```

```

: APPLICATION NUMBER: US/08/374,843B
: FILING DATE: 18-JAN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary, Kathryn Ph.D.
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 567-2020
: TELEFAX: (215) 567-2991
: TELEX: 831-494
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 992 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 200..652
: US-08-374-843B-9

Query Match          62.4%  Score 15.6;  DB 1;  Length 992;
Best local Similarity 81.8%;  Pred. No. 95;
Matches 18;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY  2 ttgtcgcaagtcacagaattat 23
      |||||  |||  |||  |||
Db  459 TTGTGCAAGTCACATAGTTAT 480

RESULT 33
US-08-905-420-9
: Sequence 9, Application US/08905420
: Patent No. 5861255
: GENERAL INFORMATION:
: APPLICANT: Demuth, Donald R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
: TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: 1601 Market Street, 36th Floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103-2398
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,420
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/374,843
: FILING DATE: 18-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary, Kathryn Ph.D.
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 567-2020
: TELEFAX: (215) 567-2991
: TELEX: 831-494
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
```

ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn Ph.D.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-374-843B-4

Query Match 62.4%; Score 15.6; DB 1; Length 450;
Best Local Similarity 81.8%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattat 23
|||||
Db 260 TTGTGCGCAAGTCGCAATGATTAT 281

RESULT 29
US-08-374-843B-18
Sequence 18, Application US/08374843B
Patent No. 5726016
GENERAL INFORMATION:
APPLICANT: Demuth, Donald R.
APPLICANT: Lally, Edward T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Street
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,843B
FILING DATE: 18-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn Ph.D.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-374-843B-18

Query Match 62.4%; Score 15.6; DB 1; Length 450;

Best Local Similarity 81.8%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 ttgtcgcaagtcacagaattat 23
|||||
Db 260 TTGTGCGCAAGTCGCAATGATTAT 281

RESULT 30
US-08-905-420-4
Sequence 4, Application US/08905420
Patent No. 5861255
GENERAL INFORMATION:
APPLICANT: Demuth, Donald R.
APPLICANT: Lally, Edward T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,420
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/374,843
APPLICATION NUMBER: US 08/374,843
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn Ph.D.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-905-420-4

Query Match 62.4%; Score 15.6; DB 2; Length 450;
Best Local Similarity 81.8%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattat 23
|||||
Db 260 TTGTGCGCAAGTCGCAATGATTAT 281

RESULT 31
US-08-905-420-18
Sequence 18, Application US/08905420
Patent No. 5861255
GENERAL INFORMATION:
APPLICANT: Demuth, Donald R.
APPLICANT: Lally, Edward T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (503) 226-7391
: TELEFAX: (503) 226-9446
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 10968 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double stranded
:   TOPOLOGY: linear
: US-08-680-327-2

Query Match          64.0% Score 16; DB 2; Length 10968;
Best Local Similarity 79.2%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25
    ||| ||||| ||| |||||
Db 745 ttgtgcgaagtcaaatgttatTT 722

RESULT 26
US-09-228-246-1/c
: Sequence 1, Application US/09228246
: Patent No. 6245510
: GENERAL INFORMATION:
: APPLICANT: Staskawicz, B. S. et al.
: TITLE OF INVENTION: PRF Protein and Nucleic Acid Sequences: Compositions
: TITLE OF INVENTION: and Methods for Plant Pathogen Resistance
: FILE REFERENCE: 51700
: CURRENT APPLICATION NUMBER: US/09/228,246
: EARLIER FILING DATE: 1999-01-11
: EARLIER APPLICATION NUMBER: 08/680,327
: EARLIER FILING DATE: 1996-07-11
: EARLIER APPLICATION NUMBER: 08/310,912
: EARLIER FILING DATE: 1994-09-22
: EARLIER APPLICATION NUMBER: 08/227,360
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 10968
: TYPE: DNA
: ORGANISM: Lycopersicon esculentum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (3879)..(8186)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (8300)..(9466)
: US-09-228-246-1

Query Match          64.0% Score 16; DB 4; Length 10968;
Best Local Similarity 79.2%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25
    ||| ||||| ||| |||||
Db 745 ttgtgcgaagtcaaatgttatTT 722

RESULT 27
US-08-786-606-2
: Sequence 2, Application US/08786606
: Patent No. 5861495
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Au-Young, Janice
: APPLICANT: Coleman, Roger
: APPLICANT: Goli, Surya K.
: TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
: TITLE OF INVENTION: PROTEINS
```

```

: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/786,606
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy RJ
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0173 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 828 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: US-08-786-606-2

Query Match          63.2% Score 15.8; DB 2; Length 828;
Best Local Similarity 89.5%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gcaagtcacagaattatt 25
    | ||||| ||||| |||
Db 90 GAAGTCACAGAAATTAATT 108

RESULT 28
US-08-374-843B-4
: Sequence 4, Application US/08374843B
: Patent No. 5726016
: GENERAL INFORMATION:
: APPLICANT: Demuth, Donald R.
: APPLICANT: Lally, Edward T.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
: TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
: TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: 1601 Market Street, 36th Street
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103-2398
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/374,843B
: FILING DATE: 18-JAN-1995
: CLASSIFICATION: 435
```


Patent No. 6265158
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
TITLE OF INVENTION: GENOMIC ORGANIZATION
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6265158thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,014C
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290,00028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 9171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 11q22-23
US-08-952-014C-2

Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaattatt 25
|| ||||| ||||| |||||
Db 1200 TCAGAAAGTCACAGAAATGATT 1220

RESULT 19
US-08-874-266-1
Sequence 1, Application US/08874266
Patent No. 5955279
GENERAL INFORMATION:
APPLICANT: Gatl, Richard A.
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA-MUTATIONS IN THE ATM GENE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson and Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/874,266
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ways Vensko, Nancy
REGISTRATION NUMBER: 36,298
REFERENCE/DOCKET NUMBER: UCLA006,006A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9385 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 190...9357
OTHER INFORMATION:
US-08-874-266-1

Query Match 64.8%; Score 16.2; DB 2; Length 9385;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaattatt 25
|| ||||| ||||| |||||
Db 1389 TCAGAAAGTCACAGAAATGATT 1409

RESULT 20
US-08-508-836A-9
Sequence 9, Application US/08508836A
Patent No. 5777093
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
TITLE OF INVENTION: Ataxia-Telangiectasia Gene
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reising, Ethington, Barnard & Perry
STREET: P.O. Box 4390
CITY: Troy
STATE: Michigan
COUNTRY: US
ZIP: 48099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,836A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: P-313 (TAU)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9870 base pairs

RESULT 15
US-08-629-001A-2
; Sequence 2, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290.00032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 11q22-23
US-08-629-001A-2
Query Match 64.8%; Score 16.2; DB 2; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
|| |||||
Db 1200 TCAGAACTCAGAGATGATT 1220
RESULT 16
US-08-642-274D-2
; Sequence 2, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
; FILE REFERENCE: 229000033
; CURRENT APPLICATION NUMBER: US/08/642,274D
; CURRENT FILING DATE: 1996-05-03
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 9171
; TYPE: DNA
; ORGANISM: Homo Sapiens

FEATURE:
; OTHER INFORMATION:
US-08-642-274D-2
Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
|| |||||
Db 1200 tcgaagtcacagaattatt 1220
RESULT 17
US-08-952-127-2
; Sequence 2, Application US/08952127
; Patent No. 6211336
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 6211336thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,127
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290.00029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 11q22-23
US-08-952-127-2
Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
|| |||||
Db 1200 TCAGAACTCAGAGATGATT 1220
RESULT 18
US-08-952-014C-2
; Sequence 2, Application US/08952014C

LENGTH: 1374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-123-761A-2

Query Match 64.8%; Score 16.2; DB 1; Length 1374;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21
||| ||||| ||||| |||||
Db 1327 ATTATCGCAAGTCATTAATT 1307

RESULT 12
US-09-166-203-48
Sequence 48, Application US/09166203A
Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowsett, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0323
CURRENT APPLICATION NUMBER: US/09/166,203A
CURRENT FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48

Query Match 64.8%; Score 16.2; DB 2; Length 1771;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21
||| ||||| ||||| |||||
Db 306 agtggagcaagtcacagaatt 326

RESULT 13
US-09-377-309-48
Sequence 48, Application US/09377309B
Patent No. 6258790
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowsett, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0390
CURRENT APPLICATION NUMBER: US/09/377,309B
CURRENT FILING DATE: 1999-08-19
EARLIER APPLICATION NUMBER: 09/166,203
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 99
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA

ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-377-309-48

Query Match 64.8%; Score 16.2; DB 4; Length 1771;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21
||| ||||| ||||| |||||
Db 306 agtggagcaagtcacagaatt 326

RESULT 14
US-08-123-761A-1/c
Sequence 1, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
TITLE OF INVENTION: A SELECTABLE MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5198 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-123-761A-1

Query Match 64.8%; Score 16.2; DB 1; Length 5198;
Best Local Similarity 85.7%; Pred. No. 62;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21
||| ||||| ||||| |||||
Db 4889 ATTATCGCAAGTCATTAATT 4869

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: G10R3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 67.2% Score 16.8; DB 2; Length 3083;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attgtcgacagtcacagaatt 20
DB 3033 attgtcgacagtcacagaatt 3014

RESULT 9
US-08-642-274D-18
Sequence 18, Application US/08642274D
Patent No. 6200749
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
TITLE OF INVENTION: SCREEN FOR A PARTIAL A-T PHENOTYPE
FILE REFERENCE: 229000033
CURRENT APPLICATION NUMBER: US/08/642,274D
CURRENT FILING DATE: 1996-05-03
NUMBER OF SEQ ID NOS: 220
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 500
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: genomic
US-08-642-274D-18

Query Match 64.8% Score 16.2; DB 4; Length 500;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatt 25
DB 245 tcgcaagtcacagaattatt 265

RESULT 10
US-08-952-014C-18
Sequence 18, Application US/08952014C
Patent No. 6265158
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6265158thwestern Hwy., Suite 410
CITY: Farmington Hills

STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,014C
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290,00028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-952-014C-18

Query Match 64.8% Score 16.2; DB 4; Length 500;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatt 25
DB 245 tcgcaagtcacagaattatt 265

RESULT 11
US-08-123-761A-2/c
Sequence 2, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match 67.2%; Score 16.8; DB 1; Length 3083;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaat 20
|||||
Db 3033 ATTGTGCAAGTCACAGACT 3014

RESULT 7
Sequence 5, Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Prety, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 67.2%; Score 16.8; DB 1; Length 3083;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaat 20
|||||
Db 3033 ATTGTGCAAGTCACAGACT 3014

RESULT 8
US-08-486-269A-5/C
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows DEMONSTRATION Version 2.0D
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/411,768B
;; FILING DATE: 31-March-95
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: CH 3124/92
;; FILING DATE: 02-OCT-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: CH 2134/93
;; FILING DATE: 15-JUL-1993
;; INFORMATION FOR SEQ. ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5872 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ORGANISM: Escherichia coli
;; STRAIN: DSM498
;; IMMEDIATE SOURCE:
;; CLONE: pBO30A15-9
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1154..2308
;; IDENTIFICATION METHOD: experimental
;; OTHER INFORMATION: /codon_start=1154
;; OTHER INFORMATION: /EC_number=2.3.1.47
;; OTHER INFORMATION: /product="KAPA synthase"
;; OTHER INFORMATION: /evidence= EXPERIMENTAL
;; OTHER INFORMATION: /gene="biol"
;; OTHER INFORMATION: /number=2
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3043..3753
;; IDENTIFICATION METHOD: experimental
;; OTHER INFORMATION: /codon_start=3043
;; OTHER INFORMATION: /EC_number=6.3.3.3
;; OTHER INFORMATION: /product="DTPB synthase"
;; OTHER INFORMATION: /evidence= EXPERIMENTAL
;; OTHER INFORMATION: /gene="biol"
;; OTHER INFORMATION: /number=4
;; FEATURE:
;; NAME/KEY: RBS
;; LOCATION: 1141..1156
;; OTHER INFORMATION: /standard_name="Vehbiobiotin synthase"
;; FEATURE:
;; NAME/KEY: RBS
;; LOCATION: 3030..3045
;; OTHER INFORMATION: /standard_name="biol RBS"
;; PUBLICATION INFORMATION:
;; DOCUMENT NUMBER: WO 87/01391 B1
;; FILING DATE: 26-AUG-1986
;; PUBLICATION DATE: 07-APR-1993
;; US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtgcgaagtcacagaattattt 25
|||||
Db 140 ATTGTGCAAGTCACAGAAATTATT 164

RESULT 5
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792

;; GENERAL INFORMATION:
;; APPLICANT: Itoku, Ohji
;; APPLICANT: Haze, Shinitiro
;; APPLICANT: Kishimoto, Jiro
;; APPLICANT: Nakahama, Kazuo
;; TITLE OF INVENTION: BIOTIN OPERON
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Wegner, Cantor, Mueller & Player
;; STREET: 1233 20th Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20036-8218
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/944,259
;; FILING DATE: 19920914
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cantor, Herbert I.
;; REGISTRATION NUMBER: 24,392
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-887-0400
;; TELEFAX: 202-835-0605
;; TELEX: 440706
;; INFORMATION FOR SEQ. ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 17 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; ORIGINAL SOURCE:
;; ORGANISM: Escherichia coli
;; US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtgcgaagtcacagaat 20
|||||
Db 17 GTGCGAAGTCACAGAAAT 1

RESULT 6
US-07-718-575-5/c
; Sequence 5, Application US/07718575
; Patent No. 5202257
;; GENERAL INFORMATION:
;; APPLICANT: Heinemann Ph.D., Stephen F.
;; APPLICANT: Boulter Ph.D., James R.
;; APPLICANT: Hollmann Ph.D., Michael MNM
;; APPLICANT: Bellier Ph.D., Bernhard MNM
;; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
;; NUMBER OF SEQUENCES: 14
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
;; STREET: 444 So. Flower St., Suite 2000
;; CITY: Los Angeles
;; STATE: California
;; COUNTRY: United States
;; ZIP: 90071-2921
;; COMPUTER READABLE FORM:

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SPRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="Involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biocA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: /aminoacids="aminoacids"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="pibiotin"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter placc"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter placc"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="biocA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter placc"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtgcagtcacagattatt 25
DB 140 attgtgcagtcacagattatt 164

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: Of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1

```


ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match      100.0%  Score 25;  DB 2;  Length 1041;
Best Local Similarity 100.0%;  Pred. No. 0.0034;
Matches 25;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
|||||
DB 24 ATTGTCGCAAGTCACAGATTATT 48

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match      100.0%  Score 25;  DB 2;  Length 1041;
Best Local Similarity 100.0%;  Pred. No. 0.0034;
Matches 25;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
|||||
DB 24 ATTGTCGCAAGTCACAGATTATT 48

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Owen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:20:06 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25
Sequence: 1 atgtgcgaatcacagaattattt 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/6C.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17	68.0	17	2	US-07-944-259-2
6	16.8	67.2	3083	1	US-07-718-575-5
7	16.8	67.2	3083	1	US-08-481-206-5
8	16.8	67.2	3083	1	US-08-486-269A-5
9	16.2	64.8	500	4	US-08-642-274D-18
10	16.2	64.8	500	4	US-08-952-014C-18
11	16.2	64.8	1374	1	US-08-123-761A-2
12	16.2	64.8	1771	4	US-09-166-203-48
13	16.2	64.8	1771	4	US-09-377-309-48
14	16.2	64.8	5198	1	US-08-123-761A-1
15	16.2	64.8	9171	2	US-08-629-001A-2
16	16.2	64.8	9171	4	US-08-642-274D-2
17	16.2	64.8	9171	4	US-08-952-127-2
18	16.2	64.8	9171	4	US-08-952-014C-2
19	16.2	64.8	9385	1	US-08-874-266-1
20	16.2	64.8	9870	1	US-08-508-836A-9
21	16.2	64.0	3442	1	US-08-529-064-3
22	16.2	64.0	3459	4	US-08-980-060-3
23	16.2	64.0	3459	4	US-09-307-195-3
24	16.2	64.0	5150	4	US-09-086-140A-14
25	16.2	64.0	10968	4	US-08-680-327-2
26	16.2	64.0	10968	4	US-09-228-246-1
27	15.8	63.2	828	2	US-08-786-606-2

28	15.6	62.4	450	1	US-08-374-843B-4	Sequence 4, Appl
29	15.6	62.4	450	1	US-08-374-843B-18	Sequence 18, Appl
30	15.6	62.4	450	2	US-08-905-420-4	Sequence 4, Appl
31	15.6	62.4	450	2	US-08-905-420-18	Sequence 18, Appl
32	15.6	62.4	992	1	US-08-374-843B-9	Sequence 9, Appl
33	15.6	62.4	992	2	US-08-905-420-9	Sequence 9, Appl
34	15.6	62.4	997	2	US-08-374-843B-16	Sequence 16, Appl
35	15.6	62.4	997	2	US-08-905-420-16	Sequence 16, Appl
36	15.6	62.4	3348	5	US-08-222-616-35	Sequence 35, Appl
37	15.6	62.4	3348	5	PCT-US95-04228-35	Sequence 35, Appl
38	15.4	61.6	946	3	US-08-916-443A-15	Sequence 15, Appl
39	15.4	61.6	1070	3	US-08-904-284-1	Sequence 1, Appl
40	15.4	61.6	1116	3	US-08-916-443A-16	Sequence 16, Appl
41	15.4	61.6	2215	2	US-08-980-329C-1	Sequence 1, Appl
42	15.4	61.6	2437	3	US-08-904-284-6	Sequence 6, Appl
43	15.4	61.6	2946	1	US-08-687-379-1	Sequence 1, Appl
44	15.4	61.6	2946	1	US-08-687-379-3	Sequence 3, Appl
45	15.4	61.6	3220	1	US-08-254-573-1	Sequence 1, Appl
46	15.4	61.6	4368	5	PCT-US95-04567-3	Sequence 3, Appl
47	15.4	61.6	5994	3	US-08-032-365A-11	Sequence 11, Appl
48	15.4	61.6	8010	4	US-09-521-526-2	Sequence 2, Appl
49	15.4	61.6	8010	5	PCT-US95-11859-2	Sequence 2, Appl
50	15.4	61.6	9620	4	US-08-952-127-11	Sequence 11, Appl
51	15.2	60.8	46	1	US-08-173-936-167	Sequence 167, App
52	15.2	60.8	46	2	US-08-475-228A-167	Sequence 167, App
53	15.2	60.8	46	3	US-08-482-060A-167	Sequence 167, App
54	15.2	60.8	46	4	PCT-US93-12388-167	Sequence 167, App
55	15.2	60.8	46	5	US-09-328-111-497	Sequence 9, Appl
56	15.2	60.8	633	4	US-08-687-379-9	Sequence 9, Appl
57	15.2	60.8	2989	1	US-08-687-379-11	Sequence 11, Appl
58	15.2	60.8	246240	2	US-08-724-394A-20	Sequence 20, Appl
59	15.2	60.8	246240	2	US-08-724-394A-21	Sequence 21, Appl
60	15.2	60.8	246240	2	US-08-724-394A-22	Sequence 22, Appl
61	15.2	60.8	246240	2	US-08-381-631-17	Sequence 17, Appl
62	15.2	60.8	2119	2	US-09-085-199B-44	Sequence 44, Appl
63	15.2	60.8	3715	2	US-08-648-298-1	Sequence 1, Appl
64	15.2	60.8	3889	2	US-08-968-685A-9	Sequence 9, Appl
65	15.2	60.8	5616	2	US-08-996-306-1	Sequence 1, Appl
66	15.2	60.8	5616	4	US-09-338-907-1	Sequence 1, Appl
67	15.2	60.8	5616	4	US-09-338-907-1	Sequence 1, Appl
68	15.2	60.8	5616	4	US-09-338-907-1	Sequence 1, Appl
69	15.2	60.8	4403765	4	US-09-103-840A-2	Sequence 2, Appl
70	15.2	60.8	4411529	4	US-09-103-840A-1	Sequence 1, Appl
71	14.8	59.2	11298	4	US-07-869-933-31	Sequence 31, Appl
72	14.8	59.2	11298	4	US-08-201-879A-2	Sequence 2, Appl
73	14.8	59.2	11298	4	US-09-103-663-31	Sequence 31, Appl
74	14.8	59.2	14507	3	US-08-785-150-1	Sequence 1, Appl
75	14.6	58.4	52	4	US-08-485-358B-15	Sequence 15, Appl
76	14.6	58.4	444	4	US-09-018-588A-19	Sequence 19, Appl
77	14.6	58.4	602	1	US-08-229-393-1	Sequence 1, Appl
78	14.6	58.4	1476	4	US-08-896-346-1	Sequence 1, Appl
79	14.6	58.4	1758	3	US-08-836-567-3	Sequence 3, Appl
80	14.6	58.4	1953	4	US-09-232-191-28	Sequence 28, Appl
81	14.6	58.4	1953	4	US-09-232-200-78	Sequence 78, Appl
82	14.6	58.4	1953	4	US-09-232-191-78	Sequence 78, Appl
83	14.6	58.4	2360	3	US-08-836-567-9	Sequence 9, Appl
84	14.6	58.4	2931	4	US-09-056-105-15	Sequence 15, Appl
85	14.6	58.4	5894	3	US-08-665-259-24	Sequence 24, Appl
86	14.6	58.4	5894	3	US-08-762-500-74	Sequence 74, Appl
87	14.6	58.4	6525	3	US-08-762-500-74	Sequence 74, Appl
88	14.4	57.6	491	4	US-09-400-208B-23	Sequence 23, Appl
89	14.4	57.6	497	4	US-09-400-208B-24	Sequence 24, Appl
90	14.4	57.6	530	4	US-08-975-762-7	Sequence 7, Appl
91	14.4	57.6	530	4	US-08-821-324-7	Sequence 7, Appl
92	14.4	57.6	530	4	US-08-925-028-7	Sequence 7, Appl
93	14.4	57.6	530	4	US-09-102-278D-38	Sequence 38, Appl
94	14.4	57.6	537	4	US-08-642-274D-38	Sequence 38, Appl
95	14.4	57.6	537	4	US-08-952-014C-38	Sequence 38, Appl
96	14.4	57.6	545	1	US-08-463-115-43	Sequence 43, Appl
97	14.4	57.6	545	1	US-08-463-388-43	Sequence 43, Appl
98	14.4	57.6	607	4	US-09-385-982A-35	Sequence 35, Appl
99	14.4	57.6	895	3	US-09-049-677A-27	Sequence 27, Appl
100	14.4	57.6	912	1	US-08-276-887A-13	Sequence 13, Appl

KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutelin; ss.

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 1..8280
 FT /tag= a
 FT /transl_except= (pos: 787..789, aa: Val)
 FT /transl_except= (pos: 1660..1662, aa: Ala)
 FT /transl_except= (pos: 1909..1911, aa: His)
 FT /transl_except= (pos: 2248..2250, aa: Asn)
 FT misc_difference 8268..8269
 FT /tag= b
 FT /note= "site of 403 nucleotide deletion"

W09636691-A1.

21-NOV-1996.

16-MAY-1996; 96WO-0507025.

08-APR-1996; 96US-0629001.

16-MAY-1995; 95US-0441822.

(KOHN/) KOHN K I.
 (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

Shiloh Y.

WPI: 1997-012070/01.

P-PSDB: AAW19652.

New isolated ataxia-telangiectasia gene - used to develop prods. for the study, diagnosis and treatment of ataxia-telangiectasia.

Claim 5; Page -: 153pp; English.

AA768731-T68785 represent mutations of the ATM gene of the invention (see AA743497 for wild type sequence). The ATM gene of the invention, is located in the human genome at chromosome segment 11q22-23. Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or deletions (such as represented by this sequence) in the coding region of the ATM gene. A-T is a progressive genetic disorder affecting the central nervous and immune systems. A-T involves chromosomal instability, cancer predisposition, radiation sensitivity, and cell cycle abnormalities. A-T is a multi-system disease inherited in an autosomal recessive manner. The wild type gene, can be used in methods for detecting carriers of a defective gene that causes A-T. The gene can also be used to generate antibodies. The methods and antibodies can be used in the study, diagnosis and therapy of A-T.

SQ Sequence 8767 BP: 2868 A; 1569 C; 1828 G; 2502 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8767;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatctt 25
 || |||||
 Db 1200 tcagaagtcacagaatgattt 1220

Search completed: December 26, 2001, 12:17:41
 Job time: 6822 sec

```

XX AA099462;
AC
XX 30-DEC-1995 (first entry)
DT
XX Maize HMI genomic DNA.
DE
XX HMI gene: Cochliobolus carbonum Nelson race 1; fungus:
KW fungal disease toxin; resistance; ss.
XX
XX Zea mays.
OS
XX WO9507989-A.
PN
XX 23-MAR-1995.
PD
XX 19-SEP-1994; 94MO-US10497.
PF
XX 17-SEP-1993; 93US-0123761.
PR
XX (PION-) PIONEER HT-BRED INT INC.
PA
XX Briggs SP, Johal GS;
PI
XX WPI; 1995-131357/17.
DR
XX Disease resistance gene Hmi from maize - is used to confer
PT resistance to Cochliobolus carbonum.
PS
XX Claim 2; Page 20-22; 26pp; English.
XX
XX A DNA sequence is claimed having 90% translational homology to
CC genomic or cDNA sequence of maize HMI gene (AA099462 & AA099463). The
CC gene HMI controls resistance to Cochliobolus carbonum Nelson race 1.
CC The HMI gene can be used in conjunction with HC-coxin in a
CC selectable marker system for use in maize transformation.
CC This gene can also be inserted into the genomes of maize varieties
CC which lack the gene to produce transformants that are resistance
CC to disease caused by C. carbonum. The HMI resistance gene is one of
CC a family of homologous disease resistance genes in maize and other
CC crops. The invention encompasses methods for identification of
CC homologous disease resistance genes by using this gene or a
CC fragment of it as a probe.
CC
XX Sequence 5198 BP; 1375 A; 1159 C; 1136 G; 1528 T; 0 other;
SQ
Query Match 64.8%; Score 16.2; DB 16; Length 5198;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 attgtcgcaagtcacagaatt 21
DB 4889 ATTATCGCAAGTCATVAAATT 4869

```

```

RESULT 44
AAT68772
ID AAT68772 standard; DNA: 8341 BP.
XX
XX AAT68772;
XX
XX 19-AUG-1997 (first entry)
DT
XX
XX ATM mutant 4777del830 coding sequence.
DE
XX
XX ATM gene: ataxia-telangiectasia; progressive genetic disorder; antibody;
KW central nervous system; immune system; chromosomal instability; therapy;
KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.
XX

```

```

OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1.4878
FT CDS
FT
FT /*tag= a
FT /transl_except= (pos: 787..789, aa: Val)
FT /transl_except= (pos: 1660..1662, aa: Ala)
FT /transl_except= (pos: 1909..1911, aa: His)
FT /transl_except= (pos: 2248..2250, aa: Asn)
FT misc_difference 4776..4777
FT /*tag= b
FT /note= "Site of 830 nucleotide deletion"
XX
XX WO9636691-A1.
PN
XX 21-NOV-1996.
PD
XX 16-MAY-1996; 96MO-US07025.
PF
XX 08-APR-1996; 96US-0629001.
PR
XX 16-MAY-1995; 95US-0441822.
XX
XX (KOHN/) KOHN K I.
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI
XX Shiloah Y;
PI
XX WPI; 1997-012070/01.
DR
XX P-PSDB; AAM19689.
XX
XX New isolated ataxia-telangiectasia gene - used to develop prods. for
PT the study, diagnosis and treatment of ataxia-telangiectasia.
PS
XX Claim 5; Page -; 153pp; English.
XX
XX AAT68731-T68785 represent mutations of the ATM gene of the invention
CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is
CC located in the human genome at chromosome segment 11q22-23.
CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC deletions (such as represented by this sequence) in the coding region of
CC the ATM gene. A-T is a progressive genetic disorder affecting the central
CC nervous and immune systems. A-T involves chromosomal instability, cancer
CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T
CC is a multi-system disease inherited in an autosomal recessive manner.
CC The wild type gene, can be used in methods for detecting carriers of a
CC defective gene that causes A-T. The gene can also be used to generate
CC antibodies. The methods and antibodies can be used in the study,
CC diagnosis and therapy of A-T.
XX
XX Sequence 8341 BP; 2707 A; 1495 C; 1752 G; 2387 T; 0 other;
SQ
Query Match 64.8%; Score 16.2; DB 18; Length 8341;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
DB 1200 tcgcaagtcacagaattatt 1220

```

```

RESULT 45
AAT68735
ID AAT68735 standard; DNA: 8767 BP.
XX
XX AAT68735;
XX
XX 18-AUG-1997 (first entry)
DT
XX
XX ATM mutant 8269del403 coding sequence.
DE
XX
XX ATM gene: ataxia-telangiectasia; progressive genetic disorder; antibody;
KW central nervous system; immune system; chromosomal instability; therapy;
KW

```


Best Local Similarity 85.7%; Pred. No. 2.2e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 attgtccaaatcacagaatt 21

Db 1327 ATTATCGCAAGTCATMAAATT 1307

RESULT 39

AAZ30602

ID AAZ30602 standard; DNA; 1771 BP.

XX AAZ30602;

DT 18-JAN-2000 (first entry)

XX Mouse Integrin alpha 4 coding sequence.

DE Human; integrin; antisense; oligonucleotide; inhibition; expression;

XX very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;

XX vascular endothelial cell; vascular endothelium; migration; inflammation;

XX atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;

XX metastasis; circulatory system; autoimmune disease; Grave's disease;

XX Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.

XX Mus sp.

OS US5968826-A.

XX 19-OCT-1999.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 attgtccaaatcacagaatt 21

Db 306 agtggagcagatcacagaatt 326

RESULT 40

AAZ10264

ID AAZ10264 standard; DNA; 1771 BP.

XX AAZ10264;

DT 24-OCT-2001 (first entry)

XX Mouse integrin alpha 4 partial gene sequence.

DE Mouse; integrin alpha 4; antisense; L20788; very late antigen 4; VLA4;

XX autoimmune disease; inflammatory disease; rheumatoid arthritis;

XX multiple sclerosis; tumor metastasis; melanoma; psoriasis;

XX allergy; Grave's disease; Hashimoto's thyroiditis;

XX systemic lupus erythematosus; allograft rejection; ds.

XX Mus musculus.

OS US6258790-B1.

XX 10-JUL-2001.

XX 19-AUG-1999; 99US-0377309.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

XX 05-OCT-1998; 98US-0166203.

Example 12; Column 77-79; 49pp; English.

The sequence (Genbank accession number U20788) is the partial gene sequence of mouse integrin alpha 4, one component of very late antigen 4 (VLA4) which is involved in various autoimmune and inflammatory diseases. The invention relates to antisense inhibitors of integrin alpha 4 which target and inhibit expression of integrin alpha 4. The antisense molecules are useful for inhibiting the expression of integrin alpha4 in human cells or tissues, treating an animal having a disease or condition associated with expression of integrin alpha4, e.g., inflammatory disease or condition, autoimmune disease or condition including rheumatoid arthritis, multiple sclerosis and tumor metastases, melanoma, psoriasis, allergy, Grave's disease, Hashimoto's thyroiditis, systemic lupus erythematosus and allograft rejection, and diseases or conditions characterized by leukocyte migration into affected tissues, preferably central nervous system tissues. The antisense molecules are also useful for reducing the levels of VLA-4 and alpha4beta7 integrin in human cells or tissues, and reducing the adherence of cells of a first type e.g., melanoma cells

Query Match 64.8%; Score 16.2; DB 20; Length 1771;

Best Local Similarity 85.7%; Pred. No. 2.3e+02;

PN WO200047771-A2.
 XX
 PD 17-AUG-2000.
 XX
 PF 11-FEB-2000; 2000WO-IB00184.
 XX
 PR 12-FEB-1999; 99US-0119917.
 PR 23-MAR-1999; 99US-0275267.
 PR 07-MAY-1999; 99US-0133200.
 XX
 PA (GEST) GENSET.
 XX
 PI Blumenfeld M, Bougueleret L, Chumakov I;
 XX WPI; 2000-571881/53.
 DR
 XX Novel biallelic markers useful for detecting conditions and genotypes
 PT associated with arachidonic acid metabolism -
 XX
 PS Claim 13; Page 343; 802pp; English.
 XX
 CC The present invention describes polynucleotides including biallelic
 CC markers derived from genes involved in arachidonic acid metabolism and
 CC from genomic regions flanking those genes. Methods from the present
 CC invention may be used to select individuals for clinical trials and
 CC predict responses to treatment with drugs. The polynucleotides may be
 CC used in hybridisation assays, sequencing assays and specific
 CC amplification assays for identifying an eicosanoid-related biallelic
 CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
 CC segment of nucleotides containing an ERBM. The polynucleotides are
 CC useful in diagnostic kits. The markers may be used to detect conditions
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to
 CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
 CC exemplification of the present invention.
 CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
 CC SNPs) in the polynucleotide sequences from the present invention have
 CC been given as their corresponding degenerate bases e.g. a polymorphic
 CC base of C or T has been given as Y.
 CC
 SQ Sequence 1001 BP; 328 A; 186 C; 179 G; 307 T; 1 other;

Query Match 64.8%; Score 16.2; DR 21; Length 1001;
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaattatt 25
 ||||| ||||| ||||| |||||
 Db 373 TCCCAATCTCACTGAATTATT 353

RESULT 35
 AAC57500/C
 ID AAC57500 standard; DNA; 1001 BP.
 XX
 AC AAC57500;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Arachidonic acid metabolism related genomic biallelic marker #134.
 XX
 KW Human; biallelic marker; arachidonic acid metabolism; genotyping;
 KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;
 KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
 KW specific amplification assay; identification; ERBM; 12-LO-RBM;
 KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200047771-A2.
 XX
 PD 17-AUG-2000.
 XX

PF 11-FEB-2000; 2000WO-IB00184.
 XX
 PR 12-FEB-1999; 99US-0119917.
 PR 23-MAR-1999; 99US-0275267.
 PR 07-MAY-1999; 99US-0133200.
 XX
 PA (GEST) GENSET.
 XX
 PI Blumenfeld M, Bougueleret L, Chumakov I;
 XX WPI; 2000-571881/53.
 DR
 XX Novel biallelic markers useful for detecting conditions and genotypes
 PT associated with arachidonic acid metabolism -
 XX
 PS Claim 13; Page 344; 802pp; English.
 XX
 CC The present invention describes polynucleotides including biallelic
 CC markers derived from genes involved in arachidonic acid metabolism and
 CC from genomic regions flanking those genes. Methods from the present
 CC invention may be used to select individuals for clinical trials and
 CC predict responses to treatment with drugs. The polynucleotides may be
 CC used in hybridisation assays, sequencing assays and specific
 CC amplification assays for identifying an eicosanoid-related biallelic
 CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
 CC segment of nucleotides containing an ERBM. The polynucleotides are
 CC useful in diagnostic kits. The markers may be used to detect conditions
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to
 CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
 CC exemplification of the present invention.
 CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
 CC SNPs) in the polynucleotide sequences from the present invention have
 CC been given as their corresponding degenerate bases e.g. a polymorphic
 CC base of C or T has been given as Y.
 CC
 SQ Sequence 1001 BP; 320 A; 192 C; 179 G; 309 T; 1 other;

Query Match 64.8%; Score 16.2; DR 21; Length 1001;
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaattatt 25
 ||||| ||||| ||||| |||||
 Db 322 TCCCAATCTCACTGAATTATT 302

RESULT 36
 AAC57945/C
 ID AAC57945 standard; DNA; 1002 BP.
 XX
 AC AAC57945;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Arachidonic acid metabolism related genomic biallelic marker #579.
 XX
 KW Human; biallelic marker; arachidonic acid metabolism; genotyping;
 KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;
 KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
 KW specific amplification assay; identification; ERBM; 12-LO-RBM;
 KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200047771-A2.
 XX
 PD 17-AUG-2000.
 XX
 PF 11-FEB-2000; 2000WO-IB00184.
 XX
 PR 12-FEB-1999; 99US-0119917.
 PR 23-MAR-1999; 99US-0275267.
 XX


```

FT      /number= 11
FT      271..500
FT      /tag= C
FT      /number= 11
XX
XX      WO9636691-A1.
XX
XX      21-NOV-1996.
XX
XX      16-MAY-1996; 96WO-US07025.
XX
XX      08-APR-1996; 96US-0629001.
XX
XX      16-MAY-1995; 95US-0441822.
XX
XX      (KOHN/) KOHN K I.
XX      (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
XX      Shlloh Y;
XX
XX      WPI; 1997-012070/01.
XX
XX      New isolated ataxia-telangiectasia gene - used to develop prods. for
XX      the study, diagnosis and treatment of ataxia-telangiectasia.
XX
XX      Claim 1; Page 59; 153pp; English.
XX
XX      AAT444-T4496 represent exons of the ATM gene of the invention.
XX      Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
XX      deletions in the coding region of the ATM gene. A-T is a progressive
XX      genetic disorder affecting the central nervous and immune systems. A-T
XX      involves chromosomal instability, cancer predisposition, radiation
XX      sensitivity, and cell cycle abnormalities. A-T is a multi-system disease
XX      inherited in an autosomal recessive manner. Cerebellar ataxia that
XX      gradually develops into general motor dysfunction is the first clinical
XX      hallmark of A-T, and results from progressive loss of Purkinje cells in
XX      the cerebellum. Oculocutaneous telangiectasia (dilation of blood
XX      vessels) develops in the bulbar conjunctiva and facial skin, and is
XX      later accompanied by graying of the hair and atrophic changes in the
XX      skin. The co-occurrence of cerebellar ataxia and telangiectases in the
XX      conjunctivae and occasionally on the facial skin (the second early
XX      hallmark of A-T) usually establishes the differential diagnosis of A-T
XX      from other cerebellar ataxias. The ATM gene, can be used in methods for
XX      detecting carriers of a defective gene that causes A-T. The gene can also
XX      be used to generate antibodies. The antibodies and methods can be used in
XX      the study, diagnosis and therapy of A-T.
XX
XX      Sequence 500 BP; 164 A; 93 C; 90 G; 153 T; 0 other;
XX
XX      Query Match      64.8%; Score 16.2; DB 18; Length 500;
XX      Best Local Similarity 85.7%; Pred. No. 1.9e+02;
XX      Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX      Oy      5 tcgcaagtcacagaattatt 25
XX      || ||||| ||||| |||||
XX      Db      245 tcagaagtcacagaatgattt 265
XX
XX      RESULT 31
XX      AAF57603
XX      ID AAF57603 standard: DNA; 500 BP.
XX
XX      AAF57603;
XX
XX      11-JUN-2001 (first entry)
XX
XX      ATM genomic fragment containing exon 11.
XX
XX      ATM; mutation: ataxia-telangiectasia; A-T; genetic counselling;
XX      cancer; chromosome 11q22-23; exon; ds.
XX
XX      Homo sapiens.
XX

```

```

PN      US6200749-B1.
XX
XX      13-MAR-2001.
XX
XX      03-MAY-1996; 96US-0642274.
XX
XX      16-MAY-1995; 95US-0441822.
XX
XX      08-APR-1996; 96US-0629001.
XX
XX      (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
XX      Shlloh Y;
XX
XX      WPI; 2001-256467/26.
XX
XX      Mutated form of the ataxia-telangiectasia gene useful for confirmation
XX      of diagnosis and screening for carriers of the disease for genetic
XX      counseling -
XX
XX      Example 5; Columns 37-38; 72pp; English.
XX
XX      The invention relates to a purified and isolated gene, designated ATM,
XX      mutations of which cause ataxia-telangiectasia (A-T). The mutation is
XX      such that point mutation, deletion, insertion or rearrangement has
XX      occurred to produce a truncated protein or such that no protein
XX      initiation occurs and therefore A-T occurs. A method is also provided
XX      for screening subjects exhibiting a partial A-T phenotype for the ATM
XX      mutant gene. The method useful for screening for A-T in a subject for use
XX      in genetic counseling. The mutated forms of the ATM gene can be detected
XX      allowing diagnosis of A-T gene carriers. Identifying the carriers either
XX      by their defective gene or by their missing defective proteins leads to
XX      earlier and more consistent diagnosis of A-T gene carriers. Carriers of
XX      the disease are more cancer-prone and/or sensitive to therapeutic
XX      applications of radiation therefore better surveillance and treatment
XX      protocols can be initiated for them. Sequences AAF5796-AAF57648 are
XX      genomic fragments containing the various exon sequences of the ATM gene.
XX
XX      Sequence 500 BP; 164 A; 93 C; 90 G; 153 T; 0 other;
XX
XX      Query Match      64.8%; Score 16.2; DB 22; Length 500;
XX      Best Local Similarity 85.7%; Pred. No. 1.9e+02;
XX      Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX      Oy      5 tcgcaagtcacagaattatt 25
XX      || ||||| ||||| |||||
XX      Db      245 tcagaagtcacagaatgattt 265
XX
XX      RESULT 32
XX      AAH41919
XX      ID AAH41919 standard: DNA; 721 BP.
XX
XX      AAH41919;
XX
XX      30-AUG-2001 (first entry)
XX
XX      Human CLASP-3 genomic DNA fragment Ref 4.1.
XX
XX      Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
XX      cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
XX      antiinflammatory; antirheumatic; dermatological; uropathic;
XX      ophthalmological; antidiabetic; nephrotropic; antihypertensive; antidiabetic;
XX      neuroprotective; antistimulant; antibacterial; antisense therapy;
XX      gene therapy; chromosome 1p31.1; ds.
XX
XX      Homo sapiens.
XX
XX      WO200142297-A2.
XX
XX      14-JUN-2001.
XX
XX      13-DEC-2000; 2000WO-US34171.
XX

```

CC oligonucleotides and compositions (1) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiallergic, hypotensive and cytoskeletal activities.
 CC The antisense oligonucleotides and (1) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasodilator peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

SO Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 40298;
 Best Local Similarity 82.6%; Pred. No. 2.3e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| |||
 Db 11514 ttgcccaagtcacagaattagt 11536

RESULT 29

AAA35189 standard; DNA; 40298 BP.

XX AAA35189;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.

KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiallergic; cytoskeletal; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

PN WO200009525-A2.

PD 24-FEB-2000.

PF 03-AUG-1999; 99WO-US17712.

PR 03-AUG-1998; 98US-0095212.

PA (UYEC-) UNIV EAST CAROLINA.

PI Nyce JW;

XX WPI; 2000-205971/18.

DR New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers

PS Disclosure: Page 1221-1231; 1343p; English.

The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiallergic, cytoskeletal and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, and then the sequences
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

SO Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 40298;
 Best Local Similarity 82.6%; Pred. No. 2.3e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| |||
 Db 11514 ttgcccaagtcacagaattagt 11536

RESULT 30

AA743451 standard; DNA; 500 BP.

XX AA743451;

DT 08-AUG-1997 (first entry)

DE A7M gene exon 11.

KW A7M gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
 KW central nervous system; immune system; chromosomal instability; therapy;
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; ss.

OS Homo sapiens.

XX Key

XX intron

XX Location/Qualifiers

XX 1..110

XX /*tag= a

XX /number= 10

XX 111..270

XX /*tag= b

FT exon

CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF1434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

CC
 XX Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 32351;
 Best Local Similarity 82.6%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| ||
 Db 10689 ttgcccaagtcacagaattagt 10711

RESULT 27
 AAA35185
 ID AAA35185 standard; DNA; 32351 BP.

XX
 AC AAA35185;

DT 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.

XX
 KW Human adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytosstatic; analgesic; impaired airway;
 KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.
 XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0055212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 XX vasoconstriction, inflammation, allergies, asthma, hypertension,
 XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
 XX cancers -

XX Disclosure: Page 1212-1219; 1343pp; English.

XX The present invention describes a new composition comprising an
 XX antisense oligonucleotide (ON) with low adenosine (up to 15%) which
 XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
 XX inflammation. The ON can have antiinflammatory, antiallergic,
 XX antiasthmatic, cytosstatic and analgesic activities. The compositions are
 XX useful for the treatment of diseases associated with inflammation,
 XX impaired airways, including lung disease and diseases whose secondary
 XX effects afflict the lungs of a subject. They can be used for treating
 XX e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic

CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

CC
 XX Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 32351;
 Best Local Similarity 82.6%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| ||
 Db 10689 ttgcccaagtcacagaattagt 10711

RESULT 28
 AAF21311
 ID AAF21311 standard; DNA; 40298 BP.

XX
 AC AAF21311;

DT 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2878.

XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytosstatic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

XX (NYCE/) NYCE J W.

XX Nyce JW;

XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
 XX trigger adenosine receptors during metabolism, useful e.g. for treating
 XX cancers and respiratory obstructions -

XX Disclosure: Page 1305-1315; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense

AC AAC76029;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORF1564 polynucleotide sequence SEQ ID NO:3167.
 XX
 KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KW vulnary; antiparatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkels RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB: AABA1820.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5: Page 2384-2386; 5507pp; English.
 XX
 CC AAC7446 to AAC7606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antirheumatic; antihypertensive; antineoplastic;
 CC antihypertensive; antineoplastic; antihypertensive; antineoplastic;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 2730 BP; 839 A; 505 C; 558 G; 826 T; 2 other;

Query Match

66.4% Score 16.6; DB 21; Length 2730;

Best Local Similarity 82.6%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 attgacgaatcacaagaattat 23
 Db 1292 attagcaatcacaagaatgat 1314
 RESULT 26
 AAF21307
 ID AAF21307 standard; DNA: 32351 BP.
 XX
 AC AAF21307;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2874.
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 KW
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-679539/66.
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure: Page 1295-1303; 1592pp; English.
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antispasmodic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or

AAH33500
 ID AAH33500 standard; cDNA; 2001 BP.
 XX
 AC AAH33500;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:556.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KM colorectal carcinoma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000MO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI; 2001-235357/24.
 DR P-PSDB; AAG74069.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1; Page 2617; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Additionally, N may be used to produce the colon cancer-associated ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 2001 BP; 677 A; 352 C; 353 G; 616 T; 3 other;
 XX

Query Match 66.4%; Score 16.6; DB 22; Length 2001;
 Best Local Similarity 82.6%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX

OY 1 attgtcgcaagtcacagaattat 23
 ||| |||| ||||| ||||| ||
 Db 865 attatagcaaatcacagaattat 887

RESULT 24
 AAH18302
 ID AAH18302 standard; cDNA; 2533 BP.
 XX
 AC AAH18302;
 XX

DT 26-JUN-2001 (first entry)
 DE Human cDNA sequence SEQ ID NO:18295.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 18295; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95693 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2533 BP; 718 A; 526 C; 535 G; 754 T; 0 other;
 XX

Query Match 66.4%; Score 16.6; DB 22; Length 2533;
 Best Local Similarity 82.6%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX

OY 1 attgtcgcaagtcacagaattat 23
 ||| |||| ||||| ||||| ||
 Db 1345 attatagcaaatcacagaattat 1367

RESULT 25
 AAC76029
 ID AAC76029 standard; cDNA; 2730 BP.
 XX


```

FT mat_peptide 233..2830
XX /*tag= C
XX
XX WO9106648-A.
XX
XX PD 16-MAY-1991.
XX
XX PF 25-OCT-1990; 90WO-US06153.
XX
XX PR 27-OCT-1989; 89US-0428116.
XX
XX PA (SALK ) SALK INST FOR BIOL STUD.
XX
XX PI Heinemann SF, Boulter JR, Hollmann M, Bettler H, Jensen JE;
XX
XX DR WPI; 1991-164197/22.
XX
XX P-PSDB; AAR1991.
XX
XX PT Glutamate receptors - used to screen for functional ligands and
XX identify and isolate further receptors
XX
XX PS Disclosure; Fig 4; 109pp; English.
XX
XX CC GluR3 cDNA was isolated from a rat forebrain cDNA using a
XX low-stringency screening protocol and a radiolabeled fragment of the
XX CC GluR1 cDNA as probe. The cDNA is deposited (ATCC 68133).
XX CC The gene and protein can be used in drug screening, to
XX CC determine whether a substance is a functional ligand for the
XX CC receptor by monitoring ion channel activity.
XX CC See also AA011849-855.
XX
XX SQ Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other:

Query Match 67.2%; Score 16.8; DB 12; Length 3083;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcgcagtcacagat 20
Db 3033 ATTCTCGCAAGTCTCAGAGT 3014

RESULT 19
AAT21899/c
ID AAT21899 standard; cDNA to mRNA; 274 BP.
XX
XX AC AAT21899;
XX
XX DT 14-AUG-1996 (first entry)
XX
XX DE Human gene signature HUMGS03441.
XX
XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO9514772-A1.
XX
XX PD 01-JUN-1995.
XX
XX PF 11-NOV-1994; 94WO-JP01916.
XX
XX PR 12-NOV-1993; 93JP-0355504.
XX
XX PA (MATS/) MATSUBARA K.
XX (OKUBO/) OKUBO K.
XX
XX PI Matsubara K, Okubo K;
XX
XX DR WPI; 1995-206931/27.

```

```

XX
XX PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX PT reflects relative abundance of corresp. mRNA in specific human
XX PT tissues
XX
XX PS Claim 1; Page 996; 2245pp; Japanese.
XX
XX CC A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX CC given in AAT19001-r26837 and which is able to hybridise to part of
XX CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX CC sequences were obtained from 3'-directed cDNA libraries prepared
XX CC from various human tissues; synthesis of cDNA was initiated from the
XX CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX CC untranslated sequence is unique to a particular mRNA species, almost
XX CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX CC is constructed so as to reflect accurately the relative abundance of
XX CC different mRNAs in the particular tissue from which it was derived.
XX CC The appearance frequency of a given GS in a cDNA library can be
XX CC determined (esp. using primers and probes derived from the GS
XX CC sequences) as a means of diagnosing abnormal cell function or for
XX CC recognising different cell types.
XX
XX SQ Sequence 274 BP; 104 A; 28 C; 52 G; 87 T; 3 other:

Query Match 66.4%; Score 16.6; DB 16; Length 274;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaatcacagatatt 24
Db 187 TTATCCCAATCACAGATGATT 165

RESULT 20
AAT19139
ID AAT19139 standard; cDNA to mRNA; 322 BP.
XX
XX AC AAT19139;
XX
XX DT 04-JUL-1996 (first entry)
XX
XX DE Human gene signature HUMGS00153.
XX
XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO9514772-A1.
XX
XX PD 01-JUN-1995.
XX
XX PF 11-NOV-1994; 94WO-JP01916.
XX
XX PR 12-NOV-1993; 93JP-0355504.
XX
XX PA (MATS/) MATSUBARA K.
XX (OKUBO/) OKUBO K.
XX
XX PI Matsubara K, Okubo K;
XX
XX DR WPI; 1995-206931/27.
XX
XX PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX PT reflects relative abundance of corresp. mRNA in specific human
XX PT tissues
XX
XX PS Claim 1; Page 312; 2245pp; Japanese.

```


DR WPI: 1999-070078/06.
 XX New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 XX
 PS Claim 1: Page 498; 641pp: English.
 XX
 CC The present sequence represents an expressed sequence tag (EST), and is
 CC a polynucleotide of the invention. The polynucleotides of the invention
 CC are all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity, haemostatic
 CC activity/inhibin activity, chemotactic/chemokinetic activity, anti-inflammatory
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammation
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 CC
 SQ Sequence 512 BP; 160 A; 104 C; 96 G; 152 T; 0 other:
 XX
 XX
 Query Match 68.0%; Score 17; DB 20; Length 512;
 Best Local Similarity 80.0%; Pred. No. 82;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 atgtcgcaagtcacagaattatt 25
 ||||| | ||||| ||||| ||
 DB 82 atgtgccagtcacagaattatt 106
 RESULT 15
 AAH94145/c
 ID AAH94145 standard; cDNA: 754 BP.
 XX
 AC AAH94145;
 XX
 DT 05-OCT-2001 (first entry)
 XX
 DE Human foetal cDNA, SEQ ID NO: 674.
 XX
 KM Human: foetal protein; cytosolic; immunosuppressive; immunostimulant;
 KM neurotrophic; neuroprotective; thrombolytic; osteopathic; anti-inflammatory;
 KM gene therapy; antisense therapy; cancer; immune disorder;
 KM growth disorder; osteoporosis; thrombolytic disorder;
 KM nervous system disorder; inflammation; expressed sequence tag; EST; ss.
 KM
 OS Homo sapiens.
 XX
 PN WO20015339-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02723.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 15-SEP-2000; 2000US-0663870.
 PR 06-NOV-2000; 2000US-0707351.
 XX
 PA (HSE-) HSEQ INC.
 XX
 PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, "Yang Y";
 PI Liu C, Asundi V, Zhou P, Werhman T;
 XX
 DR WPI: 2001-065571/50.
 XX
 PT P-PSDB: AAM06470.
 PT Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune

PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation -
 XX
 PS Claim 1: Page 430; 715pp: English.
 XX
 CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are
 CC useful in the treatment and diagnosis of diseases such as cancers,
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
 CC disorders, nervous system disorders and inflammation. The present
 CC sequence was assembled using an expressed sequence tag (EST) found
 CC to be expressed in human foetal tissue cDNA libraries as the seed.
 CC
 SQ Sequence 754 BP; 190 A; 172 C; 185 G; 207 T; 0 other:
 XX
 XX
 Query Match 68.0%; Score 17; DB 22; Length 754;
 Best Local Similarity 80.0%; Pred. No. 86;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 atgtcgcaagtcacagaattatt 25
 ||||| ||| ||||| |||||
 DB 582 ATAGTCGCAATTCAGAAATTATT 558
 RESULT 16
 AAV52308/c
 ID AAV52308 standard; DNA: 4597 BP.
 XX
 AC AAV52308;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:175.
 XX
 KM Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KM computer readable medium; vaccine; pharmaceutical composition; ds.
 KM
 OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 DR WPI: 1998-272225/24.
 XX
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 XX
 PS Claim 1: Page 1105-1107; 1409pp: English.
 XX
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any

FT intron 12206..12559
 FT /tag= aj
 FT /number= "5"
 FT exon 12560..12600
 FT /tag= ak
 FT /number= "6"
 FT intron 12601..12879
 FT /tag= al
 FT /number= "6"
 FT exon 12880..12992
 FT /tag= am
 FT /number= "7"
 FT intron 12993..13449
 FT /tag= an
 FT /number= "7"
 FT exon 13450..13552
 FT /tag= ao
 FT /number= "8"

PN W0200104315-A2.

PD 18-JAN-2001.

PF 12-JUL-2000: 2000WO-US19035.

PR 12-JUL-1999: 99US-0143364.

PA (TEXA) UNIV TEXAS A & M SYSTEM.

PA (CORR) CORNELL RES FOUND INC.

PI Giovannoni J, Tanksley S, Vrebalov J, Padmanabhan V, Ruzinskiy D;

PI White R;

DR WPI: 2001-103084/11.

XX New isolated nucleic acid sequence comprising RIN (ripening-inhibitor)
 PT or MC (macrocalyx) genes for use in genetic transformation techniques
 PT to manipulate a variety of plant characteristics
 XX
 PS Disclosure: Fig 7; 167pp; English.

XX The present sequence is tomato chromosome 5 harbouring the RIN and MC
 CC genes. The invention relates to the RIN (ripening-inhibitor) and MC
 CC (macrocalyx) genes. The RIN and MC genes are useful in controlling
 CC size, control of ripening and quality, control of sepal development and
 CC of ethylene response, and DNA markers for assisted breeding. The
 CC RIN and MC genes are used in genetic transformation techniques to
 CC manipulate a variety of plant characteristics. Hence these genes
 CC represent a valuable new tool for the creation of transgenic plants,
 CC preferably having one or more added beneficial characteristics.
 XX
 SQ Sequence 13830 BP; 4991 A; 1594 C; 1910 G; 5335 T; 0 other;

Query Match

Best Local Similarity 68.8%; Score 17.2; DB 22; Length 13830;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 gtgcgaagtcacagaattattc 25
 DB 4716 gtggaagtcacaaattattt 4737

RESULT 13

AA056307/c

ID AA056307 standard; DNA: 17 BP.

XX AA056307;

DT 13-APR-1994 (first entry)

DE Biob DNA primer.

XX BioA; BioB; promoter; biotin; operon; primer; ss.
 KW Synthetic.
 XX
 OS JP05219956-A.

PN 31-AUG-1993.

PF 14-SEP-1992: 92JP-0244792.

PR 14-SEP-1992: 92JP-0244792.

PA (SHIS) SHISEIDO CO LTD.
 PA (TAKE) TAKEDA CHEM IND LTD.

DR WPI: 1993-308323/39.

XX DNA sequence of biotin operon - has base sequence of E. coli
 PT mutated by base pair(s) compared to wild type

XX Example 1; Fig 8; 11pp; Japanese.

XX A novel DNA sequence comprises the E. coli biotin operon (BO) in which
 CC the control region of BO or the region near the bioB initiation
 CC codon is mutated by at least one base pair compared to its
 CC wild type. Two primers (AA056306-056307) are described in Example 1.
 CC A microorganism belonging to Escherichia genus, transformed by
 CC a recombinant plasmid carrying such DNA can be used for the prodn.
 CC of biotin-active substances.

XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match

Best Local Similarity 68.0%; Score 17; DB 14; Length 17;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtgcgaagtcacagaatt 20
 DB 17 GTCCGACAGTCACAGAAT 1

RESULT 14

AAV88744

ID AAV88744 standard; CDNA: 512 BP.

XX AAV88744;

DT 12-FEB-1999 (first entry)

DE EST clone HK26.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;
 XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN W09845437-A2.

PD 15-OCT-1998.

PF 10-APR-1998: 98WO-US06956.

PR 10-APR-1997: 97US-0837312.

PA (GENY) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;

PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 25-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 17.6; DB 21; Length 564;
 Best Local Similarity 83.3%; Pred. No. 43;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 25
 1 ||||| ||||| ||||| |||||
 Db 446 tgtcgcaagtcacagaattattt 469

RESULT 10
 ID AAA61269 standard; DNA; 3881 BP.
 AC AAA61269;

DT 18-OCT-2000 (first entry)
 XX

DE Human secreted protein gene 10 clone HDPCP94.

XX Human: secreted protein; fusion protein; gene therapy;
 KM protein therapy; diagnosis; tissue; cancer; tumour; AIDS;
 KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;
 KM fungal infection; immunosuppressive; ds.

OS Homo sapiens.
 XX
 PN WO200029422-A1.
 XX

PD 25-MAY-2000.

PF 09-NOV-1999; 99WO-US26409.

PR 12-NOV-1998; 98US-0108207.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ni J, Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;
 PI Birse CE, Carter KC, Komatsoulis G;

XX WPI: 2000-387729/33.

PT Novel human secreted proteins useful for diagnosing, preventing,
 PT treating and ameliorating a medical condition e.g. cardiovascular
 PT disease -

XX Claim 1: Page 233-234; 295pp; English.

XX The present sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number and the clone it was derived
 CC from are given in the descriptor line.
 CC The invention relates to 31 novel genes and their fragments (nucleic
 CC acid sequences: AAA61260-A61293; amino acid sequences A612301-B12371)

CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the 31
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC and include products for the diagnosis or treatment of cancer, tumours,
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,
 CC bacterial and fungal infection. The genes are used to generate fusion
 CC proteins by linking to the gene a human immunoglobulin portion (AAA61251)
 CC for increasing stability of the fused protein as compared to the
 CC secreted protein only.

XX Sequence 3881 BP; 1043 A; 610 C; 716 G; 1512 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 3881;
 Best Local Similarity 83.3%; Pred. No. 56;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 25
 1 ||||| ||||| ||||| |||||
 Db 2645 ttatccaaagtcacagaattattt 2668

RESULT 11
 ID AAC24233/C
 ID AAC24233 standard; cDNA; 386 BP.
 AC AAC24233;

DT 06-OCT-2000 (first entry)
 XX

DE Human secreted protein 5' EST, SEQ ID NO: 28308.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PE 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1: SEQ ID 28308; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140685.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144087.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

```

FT      /*tag= V
FT      /transl_except= AGA encodes Lys.
PN      W09321326-A.
XX      28-OCT-1993.
PD      13-APR-1993; 93WO-FR00367.
XX      13-APR-1992; 92FR-0004491.
PR      13-APR-1992; 92FR-0004491.
XX      (ORSA-) ORSAN.
PA      Kazmaier M, Lacroute F, Mignotte-Vieux C, Minet M;
XX      Pompon D;
PI      WPI: 1993-351736/44.
XX      P-PSDB; AAR3581.
DR      New DNA encoding plant NADPH cytochrome P450 reductase - cloned
XX      by functional complementation in yeast, also recombinant enzyme
PT      useful in P450 mediated bioconversion processes
PS      Claim 11; Figure 9; 79pp; French.
XX      A new method for determining whether a DNA sequence encodes an NADPH
CC      cytochrome P450 reductase involves transforming yeasts with plasmids
CC      of a total cDNA bank of plant(s). The yeasts used in the procedure
CC      are incapable of producing their own NADPH cytochrome P450
CC      reductase. They are then exposed to a cytochrome P450 inhibitor at a
CC      level which is lethal to the yeast cells but not to cells which,
CC      because of the transformation, now contain an active NADPH
CC      cytochrome P450 reductase. Surviving clones are then isolated and
CC      plasmid DNA extracted. The gene is inserted into the plasmid at a
CC      site which places it under the control of an inducible promoter.
SQ      Sequence 2114 BP; 603 A; 400 C; 542 G; 569 T; 0 other;

Query Match      74.4%; Score 18.6; DB 14; Length 2114;
Best Local Similarity 84.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 atgtgcgaagtcacgaattattt 25
       ||||| ||||| ||||| |||||
DB      1338 atgttcgaagtcacgaagtcclt 1362

RESULT 8
ID      AA201425/c
XX      AA201425 standard; DNA: 1038602 BP.
AC      AA201425;
XX      07-OCT-1999 (first entry)
DT      Complete genome sequence of Chlamydia trachomatis.
XX      Vaccine, eye disease; conventional trachoma; nonendemic trachoma;
KW      paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis;
KW      nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX      bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
OS      Chlamydia trachomatis.
XX      W09928475-A2.
PN      10-JUN-1999.
PD      27-NOV-1998; 98WO-IB01939.
XX      04-NOV-1998; 98US-0107077.
PR      28-NOV-1997; 97FR-0015041.

```

```

PR      17-DEC-1997; 97FR-0016034.
XX      (GEST ) GENSET.
XX      Griffais R;
XX      WPI: 1999-371125/31.
DR      Genome sequence of Chlamydia trachomatis
XX      Claim 1; Page 373-656; 1755pp; English.
PS      The present sequence represents the complete genome of Chlamydia
CC      trachomatis. Open reading frames (ORFs) of the genome encode
CC      polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC      against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC      be used to control growth of the microorganism. Chlamydia trachomatis is
CC      responsible for a large number of diseases, e.g. eye diseases such as
CC      conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC      conjunctivitis; genital diseases such as nongonococcal urethritis,
CC      epididymitis, cervicitis, salpingitis, perithenaritis, and inclusion
CC      pneumopathy in breast feeding infants; and venereal
CC      lymphogranulomatosis. The polypeptides of the invention may be of use in
CC      treating these diseases.
SQ      Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match      71.2%; Score 17.8; DB 20; Length 1038602;
Best Local Similarity 90.5%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      5 tgcgaagtcacgaattattt 25
       ||| ||||| ||||| |||||
DB      988106 tgcgaagtcacgaattattt 988086

RESULT 9
ID      AAC44922
XX      AAC44922 standard; DNA: 564 BP.
AC      AAC44922;
XX      18-OCT-2000 (first entry)
DT      Arabidopsis thaliana DNA fragment SEQ ID NO: 44634.
XX      Hybridisation assay; genetic mapping; gene expression control;
KW      protein identification; signal transduction pathway;
KW      metabolic pathway; promoter; termination sequence; ss.
XX      Arabidopsis thaliana.
XX      EP1033405-A2.
PN      06-SEP-2000.
PD      25-FEB-2000; 2000EP-0301439.
XX      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
XX      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
XX      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
XX      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
XX      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
XX      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
XX      23-APR-1999; 99US-0130510.
PR      23-APR-1999; 99US-0130891.

```


AC	AA062386;	
DT	16-NOV-1994	(first entry)
XX		
DE	biotin-biosynthesis genes contg. plasmid pB030A-15/9.	
XX		
KW	Biotin; expression: enterobacteria; vitamin H; synthesis;	
KW	plasmid; pB030A-15/9; biob; biof; bioc; biod; bioh;	
KW	promoter ptac; biotin synthase; KAPA synthase;	
KW	8-amino-7-oxononanoate synthase; pimeoyl-CoA; DNB synthase;	
KW	delhiobiotin synthase; DAPA synthase;	
KW	S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;	
XX	seborrhoea; dermatitis; ds.	
OS	Escherichia coli DSM458.	
XX		
PH	Key	Location/Qualifiers
FT	promoter	1..96
FT		/*tag= a
FT		/function= "promoter ptac"
FT		/evidence= EXPERIMENTAL
FT		23..28
FT	-35_signal	/*tag= b
FT		/standard_name= "promoter ptac"
FT		45..50
FT		/*tag= c
FT		/evidence= EXPERIMENTAL
FT		/standard_name= "promoter ptac"
FT		105..109
FT	-10_signal	/*tag= d
FT		/evidence= EXPERIMENTAL
FT		/standard_name= "biob RBS no. 9"
FT		117..1157
FT		/*tag= e
FT		/product= "biotin synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "biob"
FT		/number= 1
FT	RBS	1141..1146
FT		/*tag= f
FT		/standard_name= "biof RBS"
FT		1154..2311
FT		/*tag= g
FT		/EC_number= 2.3.1.47
FT		/product= "KAPA synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "biof"
FT		/number= 2
FT		/standard_name= "8-amino-7-oxononanoate synthase"
FT		2284..2288
FT	RBS	2284..2288
FT		/*tag= h
FT		/standard_name= "bioc RBS"
FT		2295..3050
FT	CDS	2295..3050
FT		/*tag= i
FT		/function= "involved in pimeoyl-CoA synthesis"
FT		/product= "protein"
FT		/gene= "bioc"
FT		/number= 3
FT		3030..3033
FT	RBS	3030..3033
FT		/*tag= j
FT		/standard_name= "biod RBS"
FT		3043..3753
FT	CDS	3043..3753
FT		/*tag= k
FT		/EC_number= 6.3.3.3
FT		/product= "DTB synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "bioid15"
FT		/number= 4
FT		/standard_name= "delhiobiotin synthase"
FT		3712..3750
FT	misc_RNA	3712..3750
FT		/*tag= l
FT		/note= "bioid15 substitution"
FT		3742..3746
FT	RBS	3742..3746

FT		/tag= m
FT	/standard_name= "bioA RBS"	
FT	3750..5039	
FT	CDS	
FT	/tag= n	
FT	/EC_number= 2.6.1.62	
FT	/product= "DAPA synthase"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "bioA"	
FT	/number= 5	
FT	/standard_name= "S-adenosyl-L-methionine: 8-amino-	
FT	7-oxononanoate aminotransferase"	
FT	RBS	
FT	5088..5093	
FT	/tag= o	
FT	/standard_name= "ORF1 RBS"	
FT	5098..5574	
FT	CDS	
FT	/tag= p	
FT	/function= "unknown, involved in biotin synthesis"	
FT	/product= "protein"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "ORF1"	
FT	/number= 6	
FT	5583..5644	
FT	terminator	
FT	/tag= q	
FT	/standard_name= "rho-independent transcriptional	
FT	terminator"	
FT	stem_loop	
FT	5583..5605	
FT	/tag= r	
PN	WO9408023-A.	
XX	14-Apr-1994.	
PD		
PF	01-OCT-1993;	93MO-EP02688.
XX		
PR	02-OCT-1992;	92CH-0003124.
XX	15-JUL-1993;	93CH-0002134.
PA	(LONZ) LONZA AG.	
PI	Birch O, Brass J, Fuhrmann M, Shaw N;	
DR	WPI: 1994-135587/16.	
P-PSDB:	AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.	
XX		
PT	Biotechnological biotin prodn. using enterobacterial biotin-gene	
XX	- providing vitamin H in high yield	
PS	Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.	
CC	The sequence is derived from plasmid pB030A-15/9 contg. the	
CC	bioB, bioF, bioC, bioD and bioG genes responsible for biosynthesis	
CC	of biotin, arranged in a transcription unit. Microorganisms	
CC	contg. these DNA fragments or plasmids may be used in the prodn.	
CC	of biotin. Biotin (vitamin H) may prevent seborrhoea, dermatitis,	
CC	loss of appetite and tiredness.	
SQ	Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;	
OY	Query Match	100.0%; Score 25; DB 15; Length 5872;
	Best Local Similarity	100.0%; Pred. NO. 0.02; Matches
	25; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	1 atgtgcgaagtcacagaattattc 25	
	140 atgtgcgaagtcacagaattattc 164	
RESULT	6	
ID	AAZ56756/c	
XX	AAZ56756 standard; cDNA; 902 BP.	
XX	AAZ56756;	

PA (NOVS) NOVARTIS FINANCE CORP.
 XX are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK495 for
 PI Lactobacillus. Insertion of bio B improves biotin yields in
 XX microorganisms which export biotin, or enables growth in media contg.
 DR little or no biotin of organisms unable to synthesise biotin for their
 DR P-PSDB; AAW73906.
 XX own use.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding diaminopelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
 XX
 XX
 Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtgcgaagtcacagaattatt 25
 DB 24 atgtgcgaagtcacagaattatt 48
 XX
 RESULT 3
 ID AAN91329 standard; DNA; 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KM E.coli Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PN GB2216530-A.
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI; 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 PS
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK495 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 XX
 XX
 Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtgcgaagtcacagaattatt 25
 DB 47 atgtgcgaagtcacagaattatt 71
 XX
 RESULT 4
 ID AAN60496 standard; DNA; 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KM Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 XX
 PN JP61149091-A.
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI; 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 XX Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 CC
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;
 XX
 XX
 Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtgcgaagtcacagaattatt 25
 DB 65 atgtgcgaagtcacagaattatt 89
 XX
 RESULT 5
 ID AAO62386 standard; DNA; 5872 BP.
 XX

85	16.2	64.8	9169	18	AA168734	ATM mutant 8283del
86	16.2	64.8	9169	18	AA168731	ATM mutant 9001del
87	16.2	64.8	9170	18	AA168762	ATM mutant 2113del
88	16.2	64.8	9171	18	AA168737	ATM open reading f
89	16.2	64.8	9171	18	AA168737	ATM gene variant G
90	16.2	64.8	9171	18	AA168737	ATM gene variant G
91	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
92	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
93	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
94	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
95	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
96	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
97	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
98	16.2	64.8	9171	18	AA168737	ATM mutant A8711G
99	16.2	64.8	9171	22	AA168737	ATM gene coding se
100	16.2	64.8	9172	18	AA168737	ATM mutant 5554ins

ALIGNMENTS

RESULT 1

AA162941 standard; DNA: 839 BP.

AC AA162941:

DT 22-OCR-2001 (first entry)

DE Human genomic DNA SEQ ID NO 269.

Human: noctropic; neuroprotective; cytoskeletal; dermatological; virologic;
 immunosuppressive; antihypertensive; anti-HIV; antibacterial; vulnary;
 antiparkinsonian; antischistosomal; antianaemic; antidiabetic; cancer;
 antihypertensive; hepatotropic; cerebroprotective; antineoplastic;
 antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 neurological disease; infection; nephrotoxic; gene therapy; vaccine;
 ds.

OS Homo sapiens.

PN WO200155449-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01346.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUL-2000; 2000US-0216880.

PR 14-AUG-2000; 2000US-0218290.

PR 01-SEP-2000; 2000US-0223443.

PR 06-SEP-2000; 2000US-0230437.

PR 08-SEP-2000; 2000US-0231243.

PR 25-SEP-2000; 2000US-0234997.

PR 13-OCT-2000; 2000US-0236367.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249214.

PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476225/51.

Novel plasma membrane associated proteins useful for diagnosing,

treating, preventing and/or prognosing disorders related to the

proteins, including cancer, immune response and neuronal disorders

Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.

The invention relates to novel genes (AA162752-AA162961) and proteins (AA162347-AA162415) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other:

Qy 1 attgtcgcaagtcacagattattt 25

Db 56 atgtcgcaagtcacagattattt 80

RESULT 2

AA162941 standard; DNA: 1041 BP.

AA162941:

DT 12-APR-1999 (first entry)

DE E. coli biotin synthetase (BioB) coding sequence.

KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;

KW biotin synthetase; biotin production; vitamin H; BioB; ss.

OS Escherichia coli.

PN US5869719-A.

PD 09-FEB-1999.

PR 30-APR-1997; 97US-0846338.

PR 30-APR-1997; 97US-0846338.

PR 08-MAR-1995; 95US-0401068.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:17:28 ; Search time 366.42 Seconds
(without alignments)

Title:	US-09-396-196F-2
Perfect score:	25
Sequence:	1 atgtgcgaagtcacgaattatt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	100 summaries

Database : N.GeneSeq_1101.*

1:	/S1D52/gcgdata/geneSeq/geneSeq/NA1980.DAT.*
2:	/S1D52/gcgdata/geneSeq/geneSeq/NA1981.DAT.*
3:	/S1D52/gcgdata/geneSeq/geneSeq/NA1982.DAT.*
4:	/S1D52/gcgdata/geneSeq/geneSeq/NA1983.DAT.*
5:	/S1D52/gcgdata/geneSeq/geneSeq/NA1984.DAT.*
6:	/S1D52/gcgdata/geneSeq/geneSeq/NA1985.DAT.*
7:	/S1D52/gcgdata/geneSeq/geneSeq/NA1987.DAT.*
8:	/S1D52/gcgdata/geneSeq/geneSeq/NA1988.DAT.*
9:	/S1D52/gcgdata/geneSeq/geneSeq/NA1989.DAT.*
10:	/S1D52/gcgdata/geneSeq/geneSeq/NA1990.DAT.*
11:	/S1D52/gcgdata/geneSeq/geneSeq/NA1991.DAT.*
12:	/S1D52/gcgdata/geneSeq/geneSeq/NA1992.DAT.*
13:	/S1D52/gcgdata/geneSeq/geneSeq/NA1993.DAT.*
14:	/S1D52/gcgdata/geneSeq/geneSeq/NA1994.DAT.*
15:	/S1D52/gcgdata/geneSeq/geneSeq/NA1995.DAT.*
16:	/S1D52/gcgdata/geneSeq/geneSeq/NA1996.DAT.*
17:	/S1D52/gcgdata/geneSeq/geneSeq/NA1997.DAT.*
18:	/S1D52/gcgdata/geneSeq/geneSeq/NA1998.DAT.*
19:	/S1D52/gcgdata/geneSeq/geneSeq/NA1999.DAT.*
20:	/S1D52/gcgdata/geneSeq/geneSeq/NA2000.DAT.*
21:	/S1D52/gcgdata/geneSeq/geneSeq/NA2001.DAT.*
22:	/S1D52/gcgdata/geneSeq/geneSeq/NA2001.DAT.*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	839	22	AA162941	Human genomic DNA
2	25	100.0	1041	20	AA001303	E. coli biotin synthase
3	25	100.0	1084	10	AA091329	E. coli B10 gene.
4	25	100.0	1121	7	AA060456	Sequence encoding Biotin-biotin synthetase
5	25	100.0	5872	15	AA062286	Human transmembrane protein
6	18.6	74.4	502	21	AA056756	Human transmembrane protein
7	18.6	74.4	2114	14	AA051336	Plant NADH cytochrome b5
8	17.8	71.2	1038602	20	AA020142	Complete genome sequence of Arabidopsis thaliana
9	17.6	70.4	565	21	AA044922	Human secreted protein
10	17.6	70.4	3881	21	AA061269	Human secreted protein
11	17.2	68.8	386	21	AA024233	Human secreted protein


```

* 33993 34908: contig of 916 bp in length
* 34909 35008: gap of 100 bp
* 35009 35898: contig of 890 bp in length
* 35899 35998: gap of 100 bp
* 35999 36691: contig of 893 bp in length
* 36692 37898: contig of 907 bp in length
* 37899 37998: gap of 100 bp
* 37999 38902: contig of 904 bp in length
* 38903 39002: gap of 100 bp
* 39003 39885: contig of 883 bp in length
* 39886 39985: gap of 100 bp
* 39986 40904: contig of 919 bp in length
* 40905 41004: gap of 100 bp
* 41005 41912: contig of 908 bp in length
* 41913 42012: gap of 100 bp
* 42013 42924: contig of 912 bp in length
* 42925 43024: gap of 100 bp
* 43025 43924: contig of 900 bp in length
* 43925 44024: gap of 100 bp
* 44025 44911: contig of 887 bp in length
* 44912 45011: gap of 100 bp
* 45012 45924: contig of 913 bp in length
* 45925 46024: gap of 100 bp
* 46025 46937: contig of 913 bp in length
* 46938 47037: gap of 100 bp
* 47038 47897: contig of 860 bp in length
* 47898 47997: gap of 100 bp
* 47998 48883: contig of 886 bp in length
* 48884 48983: gap of 100 bp
* 48984 49866: contig of 883 bp in length
* 49867 49966: gap of 100 bp
* 49967 50888: contig of 922 bp in length
* 50889 50988: gap of 100 bp
* 50989 51892: contig of 904 bp in length
* 51893 51992: gap of 100 bp
* 51993 52878: contig of 886 bp in length
* 52879 52978: gap of 100 bp
* 52979 53875: contig of 897 bp in length
* 53876 53975: gap of 100 bp
* 53976 54850: contig of 875 bp in length
* 54851 54950: gap of 100 bp
* 54951 55844: contig of 884 bp in length
* 55845 55944: gap of 100 bp
* 55945 56864: contig of 920 bp in length
* 56865 56964: gap of 100 bp
* 56965 57878: contig of 914 bp in length
* 57879 57978: gap of 100 bp
* 57979 58856: contig of 878 bp in length
* 58857 58956: gap of 100 bp
* 58957 59853: contig of 897 bp in length
* 59854 59953: gap of 100 bp
* 59954 60851: contig of 898 bp in length
* 60852 60951: gap of 100 bp
* 60952 61849: contig of 898 bp in length
* 61850 61949: gap of 100 bp
* 61950 62844: contig of 895 bp in length
* 62845 62944: gap of 100 bp
* 62945 63832: contig of 888 bp in length
* 63833 63932: gap of 100 bp
* 63933 64835: contig of 903 bp in length
* 64836 64935: gap of 100 bp
* 64936 65824: contig of 889 bp in length
* 65825 65924: gap of 100 bp
* 65925 66829: contig of 905 bp in length
* 66830 66929: gap of 100 bp
* 66930 67841: contig of 912 bp in length
* 67842 67941: gap of 100 bp
* 67942 68836: contig of 895 bp in length
* 68837 68935: gap of 100 bp
* 68937 69820: contig of 884 bp in length
* 69821 69920: gap of 100 bp
* 69921 70786: contig of 866 bp in length

```

```

* 70787 70886: gap of 100 bp
* 70887 71793: contig of 907 bp in length
* 71794 71893: gap of 100 bp
* 71894 72799: contig of 906 bp in length
Query Match 74.4% Score 18.6; DB 2; Length 86763;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 atttcgcaagtcacagaattatt 25
DB 8469 AFTCTCTCAATTCACAGTATTATT 8445

```

Search completed: December 26, 2001, 10:55:14
Job time: 1876 sec

* 66930 67841: contig of 912 bp in length
 * 67842 67941: gap of 100 bp
 * 67942 68836: contig of 895 bp in length
 * 68837 68936: gap of 100 bp
 * 68937 69820: contig of 884 bp in length
 * 69821 69920: gap of 100 bp
 * 69921 70786: contig of 866 bp in length
 * 70787 70886: gap of 100 bp
 * 70887 71793: contig of 907 bp in length
 * 71794 71893: gap of 100 bp
 * 71894 72799: contig of 906 bp in length

Query Match 74.4%; Score 18.6; DB 2; Length 86763;
 Best Local Similarity 84.0%; Pred. No. 1.5e+02;
 Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 attgcgaagtcacagatttt 25
 Db 11566 ATTCTCATTTCACAGTATTATT 11590

RESULT 45
 AC016406/c AC016406 86763 bp DNA HTG 13-JUL-2000
 LOCUS Homo sapiens clone RP11-45L18, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC016406
 AC016406.2 GI:9104557
 VERSION HTG: HNGS-PHASE0.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE 1 (bases 1 to 86763) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 REFERENCE Homo sapiens, clone RP11-45L18
 AUTHORS 2 (bases 1 to 86763)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barina, N., Becker, R., Boguslavsky, L., Boukhvalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, A.,
 Cooke, P., Dearfield, K., Dewar, K., Domino, M., Donnelly, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gargana, S., Grant, G., Jones, C., Kann, L., Karas, A.,
 Lechoczy, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McDonald, P., Marquis, N.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testave, S., Tittell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced g1:6468861.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: W1BR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L1190
 Center clone name: 45_L18

* NOTE: This record contains 87 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 869: contig of 869 bp in length
 * 870 969: gap of 100 bp
 * 970 1868: contig of 899 bp in length
 * 1869 1968: gap of 100 bp
 * 1969 2885: contig of 917 bp in length
 * 2886 2985: gap of 100 bp
 * 2986 3905: contig of 920 bp in length
 * 3906 4005: gap of 100 bp
 * 4006 4892: contig of 887 bp in length
 * 4893 4992: gap of 100 bp
 * 4993 5903: contig of 911 bp in length
 * 5904 6003: gap of 100 bp
 * 6004 6905: contig of 902 bp in length
 * 6906 7005: gap of 100 bp
 * 7006 7921: contig of 916 bp in length
 * 7922 8021: gap of 100 bp
 * 8022 8923: contig of 902 bp in length
 * 8924 9023: gap of 100 bp
 * 9024 9922: contig of 899 bp in length
 * 9923 10022: gap of 100 bp
 * 10023 10923: contig of 901 bp in length
 * 10924 11023: gap of 100 bp
 * 11024 11910: contig of 887 bp in length
 * 11911 12010: gap of 100 bp
 * 12011 12889: contig of 879 bp in length
 * 12890 12989: gap of 100 bp
 * 12990 13889: contig of 900 bp in length
 * 13890 13989: gap of 100 bp
 * 13990 14897: contig of 908 bp in length
 * 14898 14997: gap of 100 bp
 * 14998 15894: contig of 897 bp in length
 * 15895 15994: gap of 100 bp
 * 15995 16883: contig of 889 bp in length
 * 16884 16983: gap of 100 bp
 * 16984 17910: contig of 927 bp in length
 * 17911 18010: gap of 100 bp
 * 18011 18909: contig of 899 bp in length
 * 18910 19009: gap of 100 bp
 * 19010 19895: contig of 886 bp in length
 * 19896 19995: gap of 100 bp
 * 19996 20898: contig of 903 bp in length
 * 20899 20998: gap of 100 bp
 * 20999 21891: contig of 893 bp in length
 * 21892 21991: gap of 100 bp
 * 21992 22877: contig of 886 bp in length
 * 22878 22977: gap of 100 bp
 * 22978 23876: contig of 899 bp in length
 * 23877 23976: gap of 100 bp
 * 23977 24900: contig of 924 bp in length
 * 24901 25000: gap of 100 bp
 * 25001 25918: contig of 918 bp in length
 * 25919 26018: gap of 100 bp
 * 26019 26917: contig of 899 bp in length
 * 26918 27017: gap of 100 bp
 * 27018 27925: contig of 908 bp in length
 * 27926 28025: gap of 100 bp
 * 28026 28916: contig of 891 bp in length
 * 28917 29016: gap of 100 bp
 * 29017 29911: contig of 895 bp in length
 * 29912 30011: gap of 100 bp
 * 30012 30911: contig of 900 bp in length
 * 30912 31011: gap of 100 bp
 * 31012 31904: contig of 893 bp in length
 * 31905 32004: gap of 100 bp
 * 32005 32903: contig of 899 bp in length
 * 32904 33003: gap of 100 bp
 * 33004 33892: contig of 889 bp in length
 * 33893 33992: gap of 100 bp

NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 869: contig of 869 bp in length
* 870 969: gap of 100 bp
* 970 1868: contig of 899 bp in length
* 1869 1968: gap of 100 bp
* 1969 2885: contig of 917 bp in length
* 2886 2985: gap of 100 bp
* 2986 3905: contig of 920 bp in length
* 3906 4005: gap of 100 bp
* 4006 4892: contig of 887 bp in length
* 4893 4992: gap of 100 bp
* 4993 5903: contig of 911 bp in length
* 5904 6003: gap of 100 bp
* 6004 6905: contig of 902 bp in length
* 6906 7005: gap of 100 bp
* 7006 7921: contig of 916 bp in length
* 7922 8021: gap of 100 bp
* 8022 8923: contig of 902 bp in length
* 8924 9023: gap of 100 bp
* 9024 9922: contig of 899 bp in length
* 9923 10022: gap of 100 bp
* 10023 10923: contig of 901 bp in length
* 10924 11023: gap of 100 bp
* 11024 11910: contig of 887 bp in length
* 11911 12010: gap of 100 bp
* 12011 12889: contig of 879 bp in length
* 12890 12989: gap of 100 bp
* 12990 13889: contig of 900 bp in length
* 13890 13989: gap of 100 bp
* 13990 14897: contig of 908 bp in length
* 14898 14997: gap of 100 bp
* 14998 15894: contig of 897 bp in length
* 15895 15994: gap of 100 bp
* 15995 16883: contig of 889 bp in length
* 16884 16983: gap of 100 bp
* 16984 17910: contig of 927 bp in length
* 17911 18010: gap of 100 bp
* 18011 18909: contig of 899 bp in length
* 18910 19009: gap of 100 bp
* 19010 19895: contig of 886 bp in length
* 19896 19995: gap of 100 bp
* 19996 20898: contig of 903 bp in length
* 20899 20998: gap of 100 bp
* 20999 21891: contig of 893 bp in length
* 21892 21991: gap of 100 bp
* 21992 22877: contig of 886 bp in length
* 22878 22977: gap of 100 bp
* 22978 23876: contig of 899 bp in length
* 23877 23976: gap of 100 bp
* 23977 24900: contig of 924 bp in length
* 24901 25000: gap of 100 bp
* 25001 25918: contig of 918 bp in length
* 25919 26018: gap of 100 bp
* 26019 26917: contig of 899 bp in length
* 26918 27017: gap of 100 bp
* 27018 27925: contig of 908 bp in length
* 27926 28025: gap of 100 bp
* 28026 28916: contig of 891 bp in length
* 28917 29016: gap of 100 bp
* 29017 29911: contig of 895 bp in length
* 29912 30011: gap of 100 bp
* 30012 30911: contig of 900 bp in length

30912 31011: gap of 100 bp
* 31012 31904: contig of 893 bp in length
* 31905 32004: gap of 100 bp
* 32005 32903: contig of 899 bp in length
* 32904 33003: gap of 100 bp
* 33004 33892: contig of 889 bp in length
* 33893 33992: gap of 100 bp
* 33993 34908: contig of 916 bp in length
* 34909 35008: gap of 100 bp
* 35009 35898: contig of 890 bp in length
* 35899 35998: gap of 100 bp
* 35999 36891: contig of 893 bp in length
* 36892 36991: gap of 100 bp
* 36992 37898: contig of 907 bp in length
* 37899 37998: gap of 100 bp
* 37999 38902: contig of 904 bp in length
* 38903 39002: gap of 100 bp
* 39003 39885: contig of 883 bp in length
* 39886 39985: gap of 100 bp
* 39986 40904: contig of 919 bp in length
* 40905 41004: gap of 100 bp
* 41005 41912: contig of 908 bp in length
* 41913 42012: gap of 100 bp
* 42013 42924: contig of 912 bp in length
* 42925 43024: gap of 100 bp
* 43025 43924: contig of 900 bp in length
* 43925 44024: gap of 100 bp
* 44025 44911: contig of 887 bp in length
* 44912 45011: gap of 100 bp
* 45012 45924: contig of 913 bp in length
* 45925 46024: gap of 100 bp
* 46025 46937: contig of 913 bp in length
* 46938 47037: gap of 100 bp
* 47038 47897: contig of 860 bp in length
* 47898 47997: gap of 100 bp
* 47998 48883: contig of 886 bp in length
* 48884 48983: gap of 100 bp
* 48984 49866: contig of 883 bp in length
* 49867 49966: gap of 100 bp
* 49967 50888: contig of 922 bp in length
* 50889 50988: gap of 100 bp
* 50989 51892: contig of 904 bp in length
* 51893 51992: gap of 100 bp
* 51993 52878: contig of 886 bp in length
* 52879 52978: gap of 100 bp
* 52979 53875: contig of 897 bp in length
* 53876 53975: gap of 100 bp
* 53976 54850: contig of 875 bp in length
* 54851 54950: gap of 100 bp
* 54951 55844: contig of 884 bp in length
* 55845 55944: gap of 100 bp
* 55945 56864: contig of 920 bp in length
* 56865 56964: gap of 100 bp
* 56965 57878: contig of 914 bp in length
* 57879 57978: gap of 100 bp
* 57979 58856: contig of 878 bp in length
* 58857 58956: gap of 100 bp
* 58957 59853: contig of 897 bp in length
* 59854 59953: gap of 100 bp
* 59954 60851: contig of 898 bp in length
* 60852 60951: gap of 100 bp
* 60952 61849: contig of 898 bp in length
* 61850 61949: gap of 100 bp
* 61950 62844: contig of 895 bp in length
* 62845 62944: gap of 100 bp
* 62945 63832: contig of 888 bp in length
* 63833 63932: gap of 100 bp
* 63933 64835: contig of 903 bp in length
* 64836 64935: gap of 100 bp
* 64936 65824: contig of 889 bp in length
* 65825 65924: gap of 100 bp
* 65925 66829: contig of 905 bp in length
* 66830 66929: gap of 100 bp


```

* 51475 51574: gap of 100 bp
* 51575 52363: contig of 789 bp in length
* 52364 52463: gap of 100 bp
* 52464 53248: contig of 785 bp in length
* 53249 53348: gap of 100 bp
* 53349 54133: contig of 785 bp in length
* 54134 54235: gap of 100 bp
* 54236 55027: contig of 794 bp in length
* 55028 55928: contig of 801 bp in length
* 55929 56028: gap of 100 bp
* 56029 56820: contig of 792 bp in length
* 56821 56920: gap of 100 bp
* 56921 57725: contig of 805 bp in length
* 57726 57825: gap of 100 bp
* 57826 58638: contig of 813 bp in length
* 58639 58738: gap of 100 bp
* 58739 59520: contig of 782 bp in length
* 59521 59620: gap of 100 bp
* 59621 60439: contig of 819 bp in length
* 60440 60539: gap of 100 bp
* 60540 61375: contig of 835 bp in length
* 61375 61474: gap of 100 bp
* 61475 62272: contig of 798 bp in length
* 62273 62372: gap of 100 bp

Query Match      74.4%: Score 18.6; DB 2; Length 82189;
Best Local Similarity 84.0%: Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 attgcgaagtcacagaattatt 25
Db 48302 ATTCTTTCATGTCACGAATATT 48326

RESULT 43
LOCUS AT112J5 84499 bp DNA PLN 24-FEB-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T12J5 (ES5A11 project).
ACCESSION AL035522
VERSION AL035522.1 GI:4455339
KEYWORDS
SOURCE
ORGANISM
.
thale cress;
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 84499)
Beyan, M., Murphy, G., Ridley, P., Hudson, S., Hoheisel, J., Mewes, H.W.,
Mayer, K.F.X. and Schellier, C.
Unpublished
2 (bases 1 to 84499)
EU Arabidopsis sequencing project.
DIRECT SUBMISSION
Submitted (22-FEB-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
Coordinator: Mike Beyan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.beyan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosome 4 can be
viewed at: http://webserver.mips.biochem.mpg.de/proj/thal/.

FEATURES
source
location/Qualifiers
1..84499
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
misc_feature
1..49380
/note="Position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL020203; for sequence analysis and annotation please
refer to this entry"
48992..51335
/gene="T12J5.10"
/complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
/complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
/note="similarity to Human (H326) mRNA, Homo sapiens,
gb:U06631"
/codon_start=1
/product="putative protein"
/protein_id="CA836721.1"
/db_xref="GI:4455340"
/translation="MSEKARGRKSNFNGILHPEVDPFRRVGVGSSIRNFSEASEN
LVLEILYKLEKHKGCNVTFSENAEDVLISGSDRRVYLMQDIGNVKSIFSGHA
NNVEQAKFMPFSDRTLVTCADGMEDLRTETAFTELTCRSVDPRRRNDAIQLAIA
IDPNSNLFVAVGMEEARLYDIRROGGLNGFTTAAADHFCPPHILGNEDVITGLA
FSDSRLIVSYNDEFIYLFPPGMLGSGNIPSPISKSPVSKSSSSSKDENHSVS
LVKGRHNCETVGVNPFGRSPSYVSGSDGRIFVWRKGGELIRVMEADRHVNCI
EPHPIETVLASSGIESDIKWYSKAABERATLPENIELIPRRIRPMLSFTHVDDE
LFGHMDIGIDNGEDESIDDIEDNDSDTSVSGVDLDDNDSDDDMDSDDDVDMD
DSDECEIEIDNNMNMNNGSGSETVSDSISGHODDVEDDD"
/complement(48992..49360)
/gene="T12J5.10"
/number=1
/complement(49361..49761)
/number=1
/complement(49762..50493)
/gene="T12J5.10"
/number=2
/complement(50494..50814)
/number=2
/complement(50815..51066)
/gene="T12J5.10"
/number=3
/complement(51067..51206)
/number=3
/complement(51207..51335)
/gene="T12J5.10"
/number=4
/complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
/complement(51806..52129)
/gene="T12J5.20"
/number=1
/gene="T12J5.20"
/complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:039301"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CA836722.1"
/db_xref="GI:4455341"
/translation="MSEKRNLLDEAKASLDIMPYVGFADIAAKCAIDLKIPKPEAI
ENMFSPQPTLSELSAVASPSHLRRTIRFLVHOGLFKEVPTDGLATGYTNPISR
RMVITKLHGKDLMAFADONLCHSOLINAMACDARVVPKAGCQGLDGVATVYV
GGGTGTGMLVLEKEFPWIKGFNPDPPIYEVAVQVDGVNNGMFDSTIPASDAVILK
WVLDHWDGDCIKILKNCEAVLPIGVKVLVECVIGKSKVTMIABERDDKLEHVRLO
LDVVMVVRHSTGKEPTLKEMDPVLTEAGFARVEVDFDVSLLIATRS"
/number=1
/complement(52130..52610)
/number=1
/complement(52611..52939)
/gene="T12J5.20"
/number=2
/complement(52940..53064)
/number=2
/complement(53065..53389)
/gene="T12J5.20"
/number=3

```

TITLE
JOURNAL
COMMENT

Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tittell, A.,
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L6911
Center clone name: 676_O_13

NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 821: contig of 821 bp in length
* 822 921: gap of 100 bp
* 922 1723: contig of 802 bp in length
* 1724 1823: gap of 100 bp
* 1824 2617: contig of 794 bp in length
* 2618 2717: gap of 100 bp
* 2718 3511: contig of 794 bp in length
* 3512 3611: gap of 100 bp
* 3612 4426: contig of 815 bp in length
* 4427 4526: gap of 100 bp
* 4527 5321: contig of 795 bp in length
* 5322 5421: gap of 100 bp
* 5422 6197: contig of 776 bp in length
* 6198 6297: gap of 100 bp
* 6298 7108: contig of 811 bp in length
* 7109 7208: gap of 100 bp
* 7209 8040: contig of 832 bp in length
* 8041 8140: gap of 100 bp
* 8141 8968: contig of 828 bp in length
* 8969 9068: gap of 100 bp
* 9069 9887: contig of 819 bp in length
* 9888 9987: gap of 100 bp
* 9988 10807: contig of 820 bp in length
* 10808 10907: gap of 100 bp
* 10908 11703: contig of 796 bp in length
* 11704 11803: gap of 100 bp
* 11804 12613: contig of 810 bp in length
* 12614 12713: gap of 100 bp
* 12714 13508: contig of 795 bp in length
* 13509 13608: gap of 100 bp
* 13609 14414: contig of 806 bp in length
* 14415 14514: gap of 100 bp
* 14515 15312: contig of 798 bp in length
* 15313 15412: gap of 100 bp
* 15413 16207: contig of 795 bp in length
* 16208 16307: gap of 100 bp
* 16308 17086: contig of 779 bp in length
* 17087 17186: gap of 100 bp
* 17187 18005: contig of 819 bp in length
* 18006 18105: gap of 100 bp

* 18106 18928: contig of 823 bp in length
* 18929 19028: gap of 100 bp
* 19029 19847: contig of 819 bp in length
* 19848 19947: gap of 100 bp
* 19948 20753: contig of 806 bp in length
* 20754 20853: gap of 100 bp
* 20854 21657: contig of 804 bp in length
* 21658 21757: gap of 100 bp
* 21758 22551: contig of 794 bp in length
* 22552 22651: gap of 100 bp
* 22652 23474: contig of 823 bp in length
* 23475 23574: gap of 100 bp
* 23575 24369: contig of 815 bp in length
* 24369 24489: gap of 100 bp
* 24490 25281: contig of 792 bp in length
* 25282 25381: gap of 100 bp
* 25382 26168: contig of 787 bp in length
* 26169 26268: gap of 100 bp
* 26269 26927: contig of 659 bp in length
* 26928 27027: gap of 100 bp
* 27028 27845: contig of 818 bp in length
* 27846 27945: gap of 100 bp
* 27946 28772: contig of 827 bp in length
* 28773 28872: gap of 100 bp
* 28873 29688: contig of 816 bp in length
* 29689 29788: gap of 100 bp
* 29789 30607: contig of 819 bp in length
* 30608 30707: gap of 100 bp
* 30708 31491: contig of 784 bp in length
* 31492 31591: gap of 100 bp
* 31592 32396: contig of 805 bp in length
* 32397 32496: gap of 100 bp
* 32497 33264: contig of 788 bp in length
* 33265 33384: gap of 100 bp
* 33385 34191: contig of 807 bp in length
* 34192 34291: gap of 100 bp
* 34292 35087: contig of 796 bp in length
* 35088 35187: gap of 100 bp
* 35188 35999: contig of 812 bp in length
* 36000 36099: gap of 100 bp
* 36100 36944: contig of 845 bp in length
* 36945 37044: gap of 100 bp
* 37045 37862: contig of 818 bp in length
* 37863 37962: gap of 100 bp
* 37963 38791: contig of 829 bp in length
* 38792 38891: gap of 100 bp
* 38892 39719: contig of 828 bp in length
* 39720 39819: gap of 100 bp
* 39820 40615: contig of 796 bp in length
* 40616 40715: gap of 100 bp
* 40716 41525: contig of 810 bp in length
* 41526 41625: gap of 100 bp
* 41626 42421: contig of 796 bp in length
* 42422 42521: gap of 100 bp
* 42522 43314: contig of 793 bp in length
* 43315 43414: gap of 100 bp
* 43415 44203: contig of 789 bp in length
* 44204 44303: gap of 100 bp
* 44304 45119: contig of 816 bp in length
* 45120 45219: gap of 100 bp
* 45220 46020: contig of 801 bp in length
* 46021 46120: gap of 100 bp
* 46121 46891: contig of 771 bp in length
* 46892 46991: gap of 100 bp
* 46992 47789: contig of 798 bp in length
* 47790 47889: gap of 100 bp
* 47890 48705: contig of 816 bp in length
* 48706 48805: gap of 100 bp
* 48806 49633: contig of 828 bp in length
* 49634 49733: gap of 100 bp
* 49734 50556: contig of 823 bp in length
* 50557 50656: gap of 100 bp
* 50657 51474: contig of 818 bp in length

exon /gene="M4E13.20"
/number=4
5934. .6029
/gene="M4E13.20"
/number=5
6030. .6296
/gene="M4E13.20"
/number=5
6297. .6459
/gene="M4E13.20"
/number=6
9714. .9955
/gene="M4E13.30"
/number=1
9714. .10406
/gene="M4E13.30"
join(9714. .9955,10256. .10406)
/gene="M4E13.30"
/note="strong similarity to actin depolymerizing factor1,
Arabidopsis thaliana, P4TCHX:G1408471
contains Actin-depolymerizing proteins signature,
[PEASRIREKMMYATRSKGLR]"
/codon_start=1
/product="actin depolymerizing factor-like protein"
/protein_id="CAAI17762.1"
/db_xref="GI:2924508"
/translation="MTDDCKSKSPMKMKVRYVYVKLESKRYTVDKVGAAGESY
DPLASIPDDCRVAFPDVYVYDNCMSKIFPITWSPASRIREKMMYATRSKGLR
RVLDVHVELQATDPTPEKGFDKIDRAK"
9956. .10255
/gene="M4E13.30"
/number=1
10256. .10406
/gene="M4E13.30"
/number=2
11099. .11172
12742. .15036
/gene="M4E13.40"
complement(12742. .15036)
/gene="M4E13.40"
/number=1
complement(12742. .15036)
/gene="M4E13.40"
complement(12742. .15036)
/protein_id="CAAI17763.1"
/db_xref="GI:2924509"
/db_xref="SPTREMBL:O49607"
/translation="MASTFVLLFLSPFISFAASAQKTFIFRIDGSMPSIFPIH
YHYSTFAEESRIVHYVTHVHGFSFAVYTPDEADNLRRHPAVILAVEDRRRELHTR
SPDGLGIDNRGLMSDSDVYITICVPTGIMPERKSRSDNLGIPKRRGCESS
ARSPNCRNKIIGARFPAQDAAYIGINKTVEFLSRADGCHTSTTAQRHIA
FKSMSSAGVAGKAVAPARIIAKYVKWDSGCUDSDILAAFDAAVDDGVDAISIS
GGCGITISPYLDPLIAGSYGAASKGIPVSSAGNEGKSVTNMLAWVTYVASTTH
DRNPPALITAGDRIAGVSLYAGVPLNRMFPVYPGKSSASLCEMTLDPKQV
RGKIVIDRGSPPRAKGLVKKAGGVMI LANGSNGELGDALHPACAVSNGEVE
DRKAYASHPNPASTDEPGTIVGIKPAPVILASGSGRPNTLSPEILKPDILAPGV
ILAWTVAVGTGLPSDPKTFENILSGTSMACPHYSGAALIKSAHDNDPVRSSA
MMPTTINLVDSNNSLIDESTGKATPYDGSGLNGRAMNGLVYDITDNDYITLQ
SIGGPTIVITRTIPRCPTPKRSPGNLNPSTIAVEPTNRKGLVSKTVIRITANA
GOAAVVRARIESRPGVTVKPPRLVETSAVKRSYAVTVVTRNRYLSETGAVF
SVTFEDGKHVRESPIVVTOMDTL"
complement(11783. .17744)
/gene="M4E13.50"
/number=1

gene
17183..18102
/gene="MAE13.50"
complement(join(17183..17744,17840..18102))
gene
/gene="MAE13.50"
complement(join(17183..17744,17840..18102))
CDS
/gene="MAE13.50"
/note="strong similarity to myb-related protein Y49, Arabidopsis thaliana, PIR1:S58292
Contains Myb DNA-binding domain repeat signatures,
[MKREEDK1][WSLIATLRPGRDINEIKMYWHY], Aldo/keto reductase
family signatures, [LPRASGAGRGCSGRL], ATP/GTP-binding
site motif A (P-loop), [GLOGRCKS]"
/codon_start=1
/product="MYB-like protein"
/protein_id="CAA17764.1"
/db_xref="GI:2924510"
/db_xref="SPTRMBL:049608"
/translation="MGKSPCCEDKHPNKGANTKEEDDKLISYIKAHGCGCKRSLPRSK
GLQCGKSCSLRWINLRPDLDKRGNFITLEDDEL I KHSILGKSKSLATRIAPGTDIE
EIKMYWHYKRLIKRKIDPATPHRIINERTSQDSSDSSKEDPVLKTLSTGSDLEK
IANGDRIDKQREVSVERECILDELRLISPMQKHLDEENIRFGYVAKYKRSAC
FGNGCKSGNNKQCFEDSSSSYSTDISSIGYDFGLGNTRVLDVSTLEAK"
complement(17745..17839)
intron
/number=1
complement(17840..18102)
exon
/gene="MAE13.50"
/number=2
complement(join(20820..21014,21397..21499,21644..21723
21807..21889,22003..22051,22140..22205,22752..22801,
22966..23090,23242..23354))
/gene="MAE13.60"
complement(join(20820..21014,21397..21499,21644..21723,

Query Match	Similarity	Score	DB	Length
Best Local	84.0%	18.6	8	80346
Matches	21	Pred. No.	1.5e+02	
	Conservative	0	Mismatches	4
			Indels	0
			Gaps	0

RESULT	42
AC024419	
LOCUS	
DEFINITION	AC024419 82189 bp DNA HTG 28-FEB-2000
ACCESSION	AC024419 Homo sapiens chromosome 1 clone RP11-676O13 map 1, LOW-PASS
VERSION	AC024419 SEQUENCE SAMPLING.
KEYWORDS	AC024419.1 GI:7108215
SOURCE	HTG; HTGS_PHASE0.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL	Birren,B., Linton,L., Nussbaum,C. and Lander,E.
REFERENCE	1 (bases 1 to 82189)
AUTHORS	Homo sapiens chromosome 1, clone RP11-676O13
TITLE	Unpublished
JOURNAL	2 (bases 1 to 82189)
REFERENCE	Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
AUTHORS	Rod,E., Roouslaviv,I,

Homo sapiens, Metazoa; Chordata; Caniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 82189)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 1, clone RP11-676013
 Unpublished
 2 (bases 1 to 82189)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Balgwin, J., Barna, N., Bedal, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burtell, G., Campiano, A., Castle, A., Cheongel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeAngelis, K., Dewar, K., Diaz, J. S., Dodge, S., Domingo, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Haas, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J., Lander, S., Largocque, K., Lehoczy, J., Levine, R., Liou, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuirk, A., McKenna, K., McPeckers, R., Meldrum, J., Meneses, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,

```

remote similarity to RAPSIN and LGN"
/db_xref="GI:6572245.1"
/translation="AASSALSLGHWYTAIGDYPNALASHKQCVLLAKOSKDELSEAR
ELGNMGAVYIAMGDENAVQCHQHLIAKDLGNKREARAYSNLGSAYHYRNPDKA
MSYHNVLELAQLEMEKAIEMRAYAGLGHARCMQLEERAKQYHQOGLIADLMDRA
AECRASSNL"
30685..36823
/feature="LIPAZ2 repeat: matches 12..6144 of consensus"
37207..37517
/feature="AluY repeat: matches 1..309 of consensus"
complement(38024..38251)
/feature="match: GSS: Em:B14059"
39179..39267
/feature="MIR repeat: matches 34..128 of consensus"
39844..39979
/feature="MIR repeat: matches 77..212 of consensus"
41186..41330
/feature="MIR repeat: matches 105..260 of consensus"
41579..41739
/feature="AluJ/FRAM repeat: matches 148..306 of consensus"
41946..42450
/feature="LIMSA repeat: matches 4820..5325 of consensus"
42451..42751
/feature="AluSg repeat: matches 1..298 of consensus"
42752..43719
/feature="LIMSA repeat: matches 5325..6295 of consensus"
43745..45171
/feature="LIMC2 repeat: matches 4830..6324 of consensus"
45174..45466
/feature="AluSg repeat: matches 1..281 of consensus"
45529..45556
/feature="LIMC4 repeat: matches 6741..6768 of consensus"
45716..45918
/feature="MIR repeat: matches 20..212 of consensus"
45970..46013
/feature="L2 repeat: matches 2157..2203 of consensus"
46148..46460
/feature="AluSc repeat: matches 1..309 of consensus"

Query Match      74.4%: Score 18.6; DB 9; Length 77601;
Best local Similarity 84.0%; Pred No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 25
||||| |||| ||||| |||
Db 16455 ATTGCTCTAAGTCAGCAATT TTTT 16431

RESULT 41
ATM4E13 80346 bp DNA PLN 07-MAY-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone M4E13 (ESSA11
DEFINITION project).
ACCESSION AL022023
VERSION AL022023.1 GI:2924505
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 80346)
Beyan, M., Purnelle, B., Boutry, M., Goffeau, A., Hohnel, J.,
Mewes, H.W., Mayer, K.F.X. and Schueller, C.
Unpublished
2 (bases 1 to 80346)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (04-APR-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. Project
Coordinator: Mike Beyan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

```

```

FEATURES
source
E-mail: michael.beyan@bsrc.ac.uk
Location/Qualifiers
1..80346
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(join(2..154,234..572))
/gene="M4E13.10"
2..572
/gene="M4E13.10"
complement(join(2..154,234..572))
/gene="M4E13.10"
/feature="similarity to Arabidopsis thaliana AT.1.24-7,
PATCH:G1532169"
/codon_start=1
/product="EF-1 alpha-like protein (fragment)"
/protein_id="CA17760.1"
/db_xref="GI:445135"
/translation="MGFGTSSSSSSSALMLGFWAVVQVSSIGNNYTPSNYSALK
SLMLTQLELNSLVAKDVGKACGILAGLASDRSPVILLISFEGILGYGOMLV
SRITQPIPMQMCVFLCMGSGSTTMMNTAVLVICIRNFRNRGPGVSGILKGYGLSTA
ITFD"
complement(2..154)
/gene="M4E13.10"
/number=1
complement(155..233)
/number=1
complement(234..572)
/gene="M4E13.10"
/number=2
join(4678..4776,5251..5332,5481..5528,5618..5824,
5954..6029,6297..6459)
/gene="M4E13.20"
/feature="strong similarity to cyclophilin, Arabidopsis
thaliana, PATCH:G244375
contains cyclophilin-type peptidyl-prolyl cis-trans
isomerase signature & profile. [YKGFPHRILISFVIOGG]"
/codon_start=1
/product="cyclophilin-like protein"
/protein_id="CA17761.1"
/db_xref="GI:2924507"
/db_xref="SPTRMBL:049605"
/translation="MRREISFLQPRCLLLVALTIFLVPALFNGKDEKQVIEDHE
ITNRVFLVDVDDQRLGRIVTIGLYGVPRVYENFALCTGEGKTSCKPLHYGTP
FHRISGFIQGGDIHQDKSSDSITGGFFPDENFKIQHSAGVAMAMTGPDSGS
QFTITTVKASWLEGEHYVLGKVIQGMDFNFAIGAGVYSGKPRKKVVIADSGELPKD
KMDEER"
4678..4776
/gene="M4E13.20"
/number=1
4678..6459
/gene="M4E13.20"
4777..5250
/gene="M4E13.20"
/number=1
5251..5332
/gene="M4E13.20"
/number=2
5333..5480
/gene="M4E13.20"
/number=2
5481..5528
/gene="M4E13.20"
/number=3
5529..5617
/gene="M4E13.20"
/number=3
5618..5824
/gene="M4E13.20"
/number=4
5825..5953

```

COMMENT

requests: clonerequest@sanger.ac.uk
 On Apr 12, 1999 this sequence version replaced g1:445556.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the NCBI database can be found at

http://www.sanger.ac.uk/projects/c_elegans/wormpep RP6-45p1 is from the library RP6-6 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pPAC4

IMPORTANT: This sequence is not the entire insert of clone RP6-45p1. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone RP6-45p1 is at 1 in this sequence. The true left end of clone CTA-544A11 is at 77498 in this sequence.

Location/Qualifiers

1..77601

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/map="q11.22-12.3"

/clone="RP6-45p1"

/clone_lib="RPC1-6"

1..240

repeat_region /note="AluSg repeat: matches 59..295 of consensus"

244..1797 /note="L1p repeat: matches 7..1565 of consensus"

1804..2246

/note="L1M4 repeat: matches 875..1311 of consensus"

complement(2320..2760)

/note="match: GSS: Em:A0359829"

2427..2750

repeat_region /note="L2 repeat: matches 1163..1510 of consensus"

3153..3447

/note="AluJo repeat: matches 1..293 of consensus"

4563..4729

/note="MLTIF repeat: matches 96..251 of consensus"

4734..5013

/note="AluSg repeat: matches 1..280 of consensus"

5043..5138

/note="MLTIF repeat: matches 8..140 of consensus"

5637..5764

/note="MIR repeat: matches 9..140 of consensus"

6121..6268

/note="L2 repeat: matches 2550..2698 of consensus"

8024..8132

/note="MIR repeat: matches 22..152 of consensus"

8129..8443

/note="L2 repeat: matches 2427..2749 of consensus"

10180..10235

/note="L2 repeat: matches 2695..2750 of consensus"

11062..11292

/note="MIR repeat: matches 20..262 of consensus"

11626..11639

/note="MIR repeat: matches 249..262 of consensus"

repeat_region

11640..11672 /note="L2 repeat: matches 2660..2692 of consensus"

repeat_region

11673..11689 /note="MIR repeat: matches 199..250 of consensus"

repeat_region

11925..12015 /note="MIR repeat: matches 63..183 of consensus"

repeat_region

12196..12339 /note="MIR repeat: matches 73..212 of consensus"

repeat_region

12804..13058 /note="MIR repeat: matches 5..262 of consensus"

repeat_region

14554..14790 /note="MIR repeat: matches 4..251 of consensus"

repeat_region

15333..15631 /note="Alu repeat: matches 1..299 of consensus"

repeat_region

15991..16210 /note="Tigger3(Golem) repeat: matches 2805..3023 of consensus"

repeat_region

16273..16541 /note="L1M6 repeat: matches 5660..5939 of consensus"

repeat_region

16542..16705 /note="Tigger3(Golem) repeat: matches 2195..2362 of consensus"

repeat_region

16706..17031 /note="L1M6 repeat: matches 5951..6294 of consensus"

repeat_region

17152..17242 /note="L1M6 repeat: matches 6034..6121 of consensus"

repeat_region

19105..19171 /note="MIR repeat: matches 156..222 of consensus"

repeat_region

19489..19617 /note="MIR repeat: matches 58..210 of consensus"

repeat_region

20380..20419 /note="20 copies 2 mer aa 80 conserved"

repeat_region

20593..20710 /note="L2 repeat: matches 2592..2710 of consensus"

repeat_region

21342..21654 /note="Alu repeat: matches 3..310 of consensus"

repeat_region

22842..22981 /note="L2 repeat: matches 2033..2168 of consensus"

repeat_region

23099..23546 /note="MLTIF repeat: matches 1..501 of consensus"

repeat_region

23787..26374 /note="L1M1 repeat: matches 2962..5632 of consensus"

repeat_region

26375..26678 /note="AluSg repeat: matches 1..299 of consensus"

repeat_region

26679..27359 /note="L1M1 repeat: matches 5632..6300 of consensus"

repeat_region

27865..28056 /note="L2 repeat: matches 1332..1536 of consensus"

repeat_region

28064..28229 /note="L1M4 repeat: matches 2839..3022 of consensus"

repeat_region

28253..28383 /note="AluJo repeat: matches 3..134 of consensus"

repeat_region

28392..28673 /note="AluJo repeat: matches 2..301 of consensus"

repeat_region

28758..28847 /note="45 copies 2 mer ag 62 conserved"

repeat_region

28851..28930 /note="40 copies 2 mer ag 63 conserved"

repeat_region

29669..29965 /note="AluJo repeat: matches 1..303 of consensus"

gene

complement(30093..30600) /gene="CH42E1.1"

CDS

complement(<30093..>30600) /gene="CH42E1.1"

note="supported by GENES and GENSCAN. match: CDNA:

Em:U04999; match: proteins: Tr:881274 Tr:058663 Tr:055489

Tr:026309 Tr:042393 Sw:Q13702 Sw:P12672 Sw:P09108

Tr:026287 Wp:CE04472"

/codon_start=1

/evidence=not_experimental

/product="d445p1.1 (PUTATIVE partial novel protein with

[illegible]

RESULT 37
 AE007865 LOCUS 8712 bp DNA BCT 27-JUL-2001
 DEFINITION Clostridium acetobutylicum ATCC824 section 353 of 356 of the complete genome.
 ACCESSION AE007865 AE001437
 VERSION AE007865.1 GI:15026800
 KEYWORDS
 SOURCE Clostridium acetobutylicum.
 ORGANISM Clostridium acetobutylicum.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
 REFERENCE 1 (bases 1 to 8712)
 AUTHORS Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitt, J., Wolf, Y. I., Tatusov, R. L., Sabatche, F., Doucelte-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V., and Smith, D. R.
 TITLE Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
 JOURNAL J. Bacteriol. 183 (16), 4823-4838 (2001)
 MEDLINE 21359325
 PUBMED 11466286
 REFERENCES 2 (bases 1 to 8712)
 AUTHORS Childress, D., Zeng, Q., and Smith, D. R.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production, Fishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA
 FEATURES
 source location/Qualifiers
 1..8712
 /organism="Clostridium acetobutylicum"
 /strain="ATCC 824"
 /db_xref="taxon:1488"
 /db_xref="ATCC:824"
 complement(134..2959)
 /gene="CAC3696"
 complement(134..2959)
 /gene="CAC3696"
 /codon_start=1
 /transl_table=1
 /product="uncharacterized conserved membrane protein, YUEB B. subtilis homolog"
 /protein_id="AAK81617.1"
 /db_xref="GI:15026801"
 /translation="MRNKTNTKNSKSPSEKDKKKSKRLSEGVILFLVCVAGLTF
 LNNLNKHQVOVQTKPIIALVNEQDPQFENGNYNGEFLPVLVSNDRKYMVOVLS
 RSVANKAVADSGVAVTYLPQNFPHNLTLOALNPKAEIDYKVLASKSOLNKLLOD
 KIVLVLPDNTSIYKMYTASVAGVANAOTNMGNVKSQGLLSLDNIVYGFPOITN
 ESYSVSIANGKISENDSMTQESNFTNSVNLNLSSESNGLPLNLTNYEDTOR
 ITDINITNANGITNOANSNDSKSYDDQYTAAYNKAQVVOFHNTDASGMEGVYAKI
 KNQISSNTLTSSVHNDIGTIDSLTKQTTLLNEKDMYLRFPGDITSPDKPTDT
 SLETDNARSAMAKMIASTGRKDNISGTAIPKTNLNLNLSVNGDYTTLLATLVK
 NGSLTPQOKDYEAKLGLVKNYATAFISIKTATADFSAPVNNSENFQIITIVK
 ATBYTLKLQASNGLSNGSISLVKPTGTGALVNDNSVTLDNRTSLTDSKGVNTAP
 TTSPTMYVYKTVLQGAASGTVAVDNGSGGSDSSVENGLIPANAIEYAGGSOF
 GYSLLNNIDTASLITFLYGAGATYSIDKSCYKSDPEASQSIIFKMYGMDVT
 QIMRSLSDDDVQEMNKGAKIKAVTDTITTLNIAISIOKEKATLSDNLSNFEVQ
 SSDQAKYNOTMOTLNAEYKANTNNITLLOQKPKRKYNNQNDLYLCYDRGCGSIVT
 ISDMVNSTAKOEDTASAOITIKSNAEFDQWVTVTKTQNDAKKVIDTNNLSTLN
 DDFKNSQNYKNAFTLSNRTDGVKADNIFNPAKPLSTRKDTTPKLANITKRDWRW
 VVIFVIGLIGALGTWIRIKPKFEK"
 gene complement(3564..3869)
 /gene="CAC3697"
 complement(3564..3869)
 /gene="CAC3697"
 /codon_start=1
 /transl_table=1
 /product="uncharacterized, Zn-finger domain containing protein, YKXC B. subtilis homolog"
 /protein_id="AAK81618.1"
 /db_xref="GI:15026802"
 /translation="MSKLKCKACGKEIARSKTSCNCGKDHNFVRHKLITVALLS
 MYVGLLTNNATLGOSVPIVILAILAALGYFVDWVKPKKVRVSDKSTSIADSSKRL"

gene complement(4248..4697)
 /gene="CAC3698"
 complement(4248..4697)
 /gene="CAC3698"
 /codon_start=1
 /transl_table=1
 /product="Hypothetical protein, CF-44 family"
 /protein_id="AAK81619.1"
 /db_xref="GI:15026803"
 /translation="MEISNKTIEVIAKSGWEGRKIDISENVKFLSEEGFEVLESAR
 FMEPEGLKINVEKKRKDGSKHTTKCTIKETVGLNSFNGLEDYVDEKVMILGSL
 YDFNLYLITSGRFSDSTAWGDSVNAEPNITINKEGTIVMKFKE"
 complement(4709..5317)
 /gene="CAC3699"
 /note="possible gene fragment 'Hypothetical protein'"
 complement(5451..5930)
 /gene="CAC3700"
 complement(5451..5930)
 /gene="CAC3700"
 /codon_start=1
 /transl_table=1
 /product="Hypothetical protein, CF-44 family"
 /protein_id="AAK81620.1"
 /db_xref="GI:15026804"
 /translation="MKETITFLIKSGHWPNRKIDITDLYVYERKGEIPEPAKKFL
 EEEGMIDVYCPINPPIPEEDIKRYFNRYDITTMKISLNGMLSRQISIEYEEYV
 EKLTVVGSINGQVLYMISECKMFEHGFENNABEEDWRLDNTDITNMQWDGFI"
 complement(6812..7309)
 /gene="CAC3701"
 complement(6812..7309)
 /gene="CAC3701"
 /codon_start=1
 /transl_table=1
 /product="Hypothetical protein, CF-44 family"
 /protein_id="AAK81621.1"
 /db_xref="GI:15026805"
 /translation="MLEERVKKILKRSQMIENOKIDITNQVKMLESEGEYFEARKP
 MEPEGVHITAKYIDSPGDEYDEHSTCYEDMKYVEHNVADEKYGERTLPCKLYG
 GEYITISSEKGFYSEGMADSDSNFNNGLLGEYKSGFLMWIDYKAKKRRQSKYK
 NEEYF"
 complement(7311..8586)
 /gene="CAC3702"
 complement(7311..8586)
 /gene="CAC3702"
 /codon_start=1
 /transl_table=1
 /product="Hypothetical protein, CF-7 family"
 /protein_id="AAK81622.1"
 /db_xref="GI:15026806"
 /translation="MLFIVNQGDIGAAAGKASKAESLLGKVGTFPSKSIASVENAK
 NJATLPMPTVDNIGSVADLRVTFSEGGALIGDLTKSVYSSAKNAENITLIGRTLD
 DYKKAQFVNGKLLPAGSEATKDPGKATLKNGKDLKGLRQVEVMPGVRVSPNI
 ESEVYKKNKKNVQDKDNFVAVKSEPKVGNPGBQBNYSEEMVYANLKEKYGEDN
 YLCTDBKSLINSVEKLDSEFRKNTINGDYKTAAYVAGENDLPVSNMANSNSHOYNTIL
 TQGEFEDKMYNKKLLKEGDESAVASKYIEHTLTDYIKTLDDEYFISKGLYKDDVY
 EMKKLDAIENKRNENRATGKFSKLTGPESEFDWNVENCACAEVMSYTRKAILNGAKDDI
 SEKCVATGSGNAYKPCENCKRFRPLNNVGVKX"
 BASE COUNT 2602 a 1702 c 1068 g 3340 t
 ORIGIN
 Query Match 74.4%; Score 18.6; DB 1; Length 8712;
 Best local Similarity 84.0%; Pred. No. 1.4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 atgtgcgaagtcaagaattatt 25
 ||| || ||||| ||||| |||||
 Db 1310 ATTATCCAAGTACAGATTATCT 1334
 RESULT 38
 CEF19C6 36493 bp DNA INV 20-JUN-2001
 LOCUS


```

repeat_region 44537..44698
/note="MIR repeat: matches 53..212 of consensus"
repeat_region 45920..45967
/note="24 copies 2 mer ac 79% conserved"
repeat_region 47253..47427
/note="MIR repeat: matches 53..248 of consensus"
repeat_region 48266..48374
/note="MER86 repeat: matches 26..143 of consensus"
misc_feature 48733..49291
/note="match: GSS: Em:A0386870
match: STS: Em:G58335"
misc_feature 48852..49191
/note="match: STS: Em:G63817"
repeat_region 48856..48978
/note="LIMC repeat: matches 1804..1927 of consensus"
repeat_region 49342..49517
/note="LIMC repeat: matches 2418..2237 of consensus"
repeat_region 49798..50003
/note="MER58C repeat: matches 1..89 of consensus"
repeat_region 50045..50241
/note="L2 repeat: matches 2553..2744 of consensus"
repeat_region 50293..50614
/note="MER2 repeat: matches 1..345 of consensus"
repeat_region 50615..50795
/note="LIM4 repeat: matches 2881..3050 of consensus"
repeat_region 50796..51092
/note="AluJo repeat: matches 1..294 of consensus"

Query Match 75.2%: Score 18.8; DB 9; Length 84478;
Best Local Similarity 90.9%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
||||| |||||||||
Db 64614 TTGTCCAAAGTCACAGATTAT 64635

RESULT 33
AC012380 108623 bp DNA PRI 26-OCT-1999
LOCUS AC012380 Genomic Sequence For Homo sapiens Clone 125H5, Chromosome 20,
DEFINITION complete sequence.
ACCESSION AC012380
VERSION AC012380.1 GI:6119509
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108623)
Spiegel, L.A., Nascimento, L.U., de la Bastide, M., Vill, D.M.,
Huang, E.N., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Shekhar, M., Schutz, K., See, L.H., Swaby, I.,
Habermann, K., Dedhia, N.N. and McComble, W.R.
Genomic Sequence For Homo sapiens Clone 125H5, Chromosome 20,
Complete Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108623)
McComble, W.R.
Direct Submission
JOURNAL Submitted (26-OCT-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
FEATURES
source
1..108623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/misc_feature 56550..56820
/note="The assembly is single stranded with single
chemistry."
BASE COUNT 33048 a 19822 c 21292 g 34461 t

```

```

ORIGIN
Query Match 75.2%: Score 18.8; DB 9; Length 108623;
Best Local Similarity 90.9%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
||||| |||||||||
Db 39442 TTGTCCAAAGTCACAGATTAT 39463

RESULT 34
AC064870 194865 bp DNA HTG 09-AUG-2001
LOCUS AC064870/C Homo sapiens chromosome 2 clone RPL1-575E20, WORKING DRAFT
DEFINITION SEQUENCE, 2 unordered pieces.
ACCESSION AC064870.4 GI:15145272
VERSION AC064870.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 194865)
Waterston, R.H.
The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 194865)
Waterston, R.H.
Direct Submission
JOURNAL Submitted (23-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 9, 2001 this sequence version replaced gi:15022763.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0575E20
----- Summary Statistics -----
Sequencing vector: M13; 26%
Sequencing vector: plasmid; 72%
Chemistry: Dye-primer ET; 25% of reads
Chemistry: Dye-terminator Big Dye; 73% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 194615 bases at least Q40
Consensus quality: 194615 bases at least Q30
Consensus quality: 194615 bases at least Q20
Insert size: 208000; agarose-fp
Insert size: 194615; sum-of-contigs
Quality coverage: 14.96 in Q20 bases; agarose-fp
Quality coverage: 15.53 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 150: contig of 150 bp in length
* 151 250: gap of unknown length
* 251 194865: contig of 194615 bp in length.
Location/Qualifiers
1..194865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
FEATURES
source

```

http://www.sanger.ac.uk/Projects/C_elegans/wormp/ This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/CGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-108A16 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-106901 is at 84379 in this sequence. The true right end of clone RP4-79086 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-108A16 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

Location/Qualifiers

1..84478

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/clone_lib="RP11-108A16"

/clone_1lb="RPI1-11.1"

441..902

/note="MULTI repeat: matches 54..568 of consensus"

complement(863..1274)

/note="match: GSS: Em:AQ409299"

complement(974..1162)

/note="match: GSS: Em:AQ413763"

1293..1648

/note="match: GSS: Em:AQ491347"

1359..1672

/note="match: GSS: Em:AQ089882"

1405..1705

/note="Aluub repeat: matches 1..305 of consensus"

2354..2542

/note="Tigerr2a repeat: matches 1..196 of consensus"

2543..2833

/note="Alusp repeat: matches 1..288 of consensus"

2834..3094

/note="Tigerr2a repeat: matches 196..434 of consensus"

complement(4171..4768)

/note="match: GSS: Em:AQ418133"

complement(4337..4768)

/note="match: GSS: Em:AQ237510"

4363..4672

/note="Alusc repeat: matches 1..305 of consensus"

complement(4575..4753)

/note="match: GSS: Em:AQ419566"

6427..6488

/note="31 copies 2 mer tg 71% conserved"

7412..7723

/note="Alusq repeat: matches 1..311 of consensus"

9182..9303

/note="MER3 repeat: matches 75..209 of consensus"

9542..9863

/note="MER33 repeat: matches -9..321 of consensus"

11769..11855

/note="MER87b repeat: matches 948..1037 of consensus"

11871..12581

/note="L1MB6 repeat: matches 5446..6164 of consensus"

12143..12710

/note="match: GSS: Em:AQ308637"

complement(12249..12756)

/note="match: GSS: Em:AQ118115"

12382..13258

/note="MER97c repeat: matches 10..777 of consensus"

13541..14065

misc_feature

/note="MULTI repeat: matches 13..541 of consensus"

complement(15380..15816)

/note="match: GSS: Em:AQ031863"

complement(15526..15839)

/note="match: STS: Em:HSPF15A1"

16822..16939

/note="Single clone region. Assembly consistent with restriction digest."

17530..17567

/note="MADE1 repeat: matches 1..38 of consensus"

18129..18871

/note="L1M4 repeat: matches 2200..2928 of consensus"

18937..19474

/note="L1M4 repeat: matches 4674..5232 of consensus"

20088..20386

/note="AluY repeat: matches 1..303 of consensus"

20424..20500

/note="Alu repeat: matches 1..77 of consensus"

20501..23893

/note="L1PA2 repeat: matches 2740..6144 of consensus"

25257..25937

/note="match: GSS: Em:A2305856"

27252..27589

/note="match: GSS: Em:A2318983"

28250..28413

/note="L2 repeat: matches 2538..2705 of consensus"

28705..29775

/note="L1PA2 repeat: matches 5073..6144 of consensus"

30032..30250

/note="MER30 repeat: matches 2..223 of consensus"

30334..30433

/note="Aluub repeat: matches 1..104 of consensus"

complement(32162..32669)

/note="match: STS: Em:G58112"

match: GSS: Em:AQ385474"

33084..33309

/note="match: GSS: Em:AQ595367 Em:AQ669389"

33877..34227

/note="LTR7 repeat: matches 1..448 of consensus"

34228..34607

/note="HERVH repeat: matches 7324..7713 of consensus"

34608..34844

/note="HERVH repeat: matches 5588..5819 of consensus"

34825..35487

/note="HERVH repeat: matches 4488..5149 of consensus"

35490..36344

/note="HERVH repeat: matches 3140..3993 of consensus"

36324..39331

/note="HERVH repeat: matches 1..2995 of consensus"

37253..37773

/note="Sequence from clone PCR. Assembly is consistent with restriction digest."

complement(37712..37773)

/note="Weak data."

37791..37829

/note="Single clone region. Assembly consistent with restriction digest."

39334..39684

/note="LTR7 repeat: matches 1..448 of consensus"

39779..40059

/note="L1ME3A repeat: matches 5836..6115 of consensus"

40657..41016

/note="LTR1A1 repeat: matches 1..365 of consensus"

4114..41167

/note="LTR repeat: matches 13..66 of consensus"

41504..41906

/note="L2 repeat: matches 2248..2667 of consensus"

43254..43645

/note="LTR1A2 repeat: matches 4..374 of consensus"

43865..44012

/note="LTR33 repeat: matches 76..256 of consensus"

44137..44198

/note="LTR33 repeat: matches 431..493 of consensus"


```

repeat_region      /rpt_family="L1MC/D"      complement(1875..2094)
repeat_region      /rpt_family="AluJo"      complement(2128..2192)
repeat_region      /rpt_family="L1R16A"      complement(2387..2408)
repeat_region      /rpt_family="AT_rich"      complement(3195..3512)
repeat_region      /rpt_family="L1"      complement(3517..3660)
repeat_region      /rpt_family="L1"      complement(3701..5580)
repeat_region      /rpt_family="L1MA"      complement(5568..6616)
repeat_region      /rpt_family="L1MA"      complement(6596..8683)
repeat_region      /rpt_family="MER63B"      complement(9112..9218)
repeat_region      /rpt_family="MSTB"      9593..9637
repeat_region      /rpt_family="(CAAAA)n"      10874..11256
repeat_region      /rpt_family="L1ME3A"      complement(14179..14461)
repeat_region      /rpt_family="AluSx"      15013..15215
repeat_region      /rpt_family="L1MAA"      15643..15944
repeat_region      /rpt_family="AluSx"      15945..15969
repeat_region      /rpt_family="(CAAAA)n"      complement(16603..16676)
repeat_region      /rpt_family="(TAAA)n"      17273..17299
repeat_region      /rpt_family="AT_rich"      complement(18248..18399)
repeat_region      /rpt_family="MIR"      19474..19532
repeat_region      /rpt_family="L1MA5"      complement(19740..19815)
repeat_region      /rpt_family="MER20"      complement(19876..20365)
repeat_region      /rpt_family="L2"      21517..21886
repeat_region      /rpt_family="L1R16C"      22416..22529
repeat_region      /rpt_family="L2"      24462..24762
repeat_region      /rpt_family="AluSg"      25169..25201
repeat_region      /rpt_family="AT_rich"      complement(25230..25495)
repeat_region      /rpt_family="L1R16C"      26138..26169
repeat_region      /rpt_family="POLY_A"      complement(26308..26716)
repeat_region      /rpt_family="MSTA"      complement(27493..27516)
repeat_region      /rpt_family="(TAAA)n"      complement(27517..27816)
repeat_region      /rpt_family="AluSx"      complement(30094..30125)
repeat_region      /rpt_family="(CA)n"      30776..30875
repeat_region      /rpt_family="MIR"      complement(31703..31768)
repeat_region      /rpt_family="(CAT)n"      complement(31892..32060)
repeat_region      /rpt_family="MER58B"      complement(32351..32597)
repeat_region      /rpt_family="MIR"      32697..32765
repeat_region      /rpt_family="L2"

repeat_region      complement(33799..33856)
repeat_region      /rpt_family="AT_rich"      34422..34460
repeat_region      /rpt_family="AT_rich"      34912..35021
repeat_region      /rpt_family="MIR"      35194..35256
repeat_region      /rpt_family="(CA)n"      35293..35749
repeat_region      /rpt_family="MLT1C"      36048..36592
repeat_region      /rpt_family="TIGGER2"      36594..36883
repeat_region      /rpt_family="AluY"      36884..37105
repeat_region      /rpt_family="TIGGER2"      37160..37408
repeat_region      /rpt_family="TIGGER2"      complement(37657..38294)
repeat_region      /rpt_family="L1M4"      38710..38755
repeat_region      /rpt_family="L1M4"      38755..39465
repeat_region      /rpt_family="L1ME"      complement(39603..40801)
repeat_region      /rpt_family="L1M4"      40801..40852
repeat_region      /rpt_family="L1"      41310..41664
repeat_region      /rpt_family="THE1B"      42024..42058
repeat_region      /rpt_family="AT_rich"      43394..43535
repeat_region      /rpt_family="MIR"      complement(43827..43964)
repeat_region      /rpt_family="MIR"      44283..44581
repeat_region      /rpt_family="L1PA7"      complement(44665..44997)
repeat_region      /rpt_family="L1ME1"      complement(44901..45198)
repeat_region      /note="Region: T02862.FBI3A8.Petal brain, Stratogene Homo sapiens cDNA clone FBI3A8 3'end"
misc_feature
repeat_region      complement(45177..45247)
repeat_region      /rpt_family="MERSA"      45492..45527
repeat_region      /rpt_family="AT_rich"      45645..46308
repeat_region      /rpt_family="L1PA2"      46311..46350
repeat_region      /rpt_family="AT_rich"      46443..46660
repeat_region      /rpt_family="L1"      46658..46731
repeat_region      /rpt_family="L1PA4"      47822..48274
repeat_region      /rpt_family="L1PA15"      49737..49813
repeat_region      /rpt_family="(TAA)n"      49856..49910
repeat_region      /rpt_family="(CA)n"      complement(50429..50548)
repeat_region      /rpt_family="L2"      51514..51543
repeat_region      /rpt_family="AT_rich"

Query Match      76.8%      Score 19.2:      DB 9:      Length 207957;
Best Local Similarity      87.5%      Pred. No.75;
Matches      21:      Conservative      0:      Mismatches      3:      Indels      0:      Gaps      0:
QY      2      ttgtcgcaagtcacagaattattt 25
Db      85514      TTGTGCAAAATTCACAGAAATTTT 85491

```



```

misc_feature      /note="assembly_fragment"
                  6456..7718
                  /note="assembly_fragment"
misc_feature      7819..9644
                  /note="assembly_fragment"
misc_feature      9745..11368
                  /note="assembly_fragment"
misc_feature      11469..13195
                  /note="assembly_fragment"
misc_feature      13296..14769
                  /note="assembly_fragment"

Query Match      76.8%; Score 19.2; DB 2; Length 194718;
Best Local Similarity 87.5%; Pred. No. 75;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 ttgcgaagcacagattattt 25
Db 95673 TTGTCACAAATGACAGAAATATT 95696

```

```

RESULT 29
AC020772
LOCUS      AC020772 198759 bp DNA HTG 05-APR-2000
DEFINITION Homo sapiens clone RP11-710K17, WORKING DRAFT SEQUENCE, 29
unordered pieces.
ACCESSION  AC020772
VERSION     AC020772.3 GI:7417754
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 198759)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 198759)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavsky,L., Bouhgalter,B., Brown,A., Burkett,C., Castle,A.,
            Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Develiano,K., Dewar,K., Domino,M., Doyle,M., Fenesfor,J.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Gallagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marcquis,N., McEwan,P., McClure,A., McKenna,K.,
            McPheeters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
            Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rotman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Turrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (09-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Apr 5, 2000 this sequence version replaced gi:850459.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: 14348
            Center clone name: 710_K_17
            ----- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: Phrap, version 0.960731
Consensus quality: 181544 bases at least Q40
Consensus quality: 189586 bases at least Q30
Consensus quality: 193248 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 195959; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1036: contig of 1036 bp in length
* 1037 1136: gap of 100 bp
* 1137 2371: contig of 1235 bp in length
* 2372 2471: gap of 100 bp
* 2472 3036: contig of 565 bp in length
* 3037 3136: gap of 100 bp
* 3137 5321: contig of 2185 bp in length
* 5322 5421: gap of 100 bp
* 5422 7352: contig of 1931 bp in length
* 7353 7452: gap of 100 bp
* 7453 9745: contig of 2293 bp in length
* 9746 9845: gap of 100 bp
* 9846 11851: contig of 2006 bp in length
* 11852 11951: gap of 100 bp
* 11952 13770: contig of 1819 bp in length
* 13771 13870: gap of 100 bp
* 13871 17767: contig of 3897 bp in length
* 17768 17867: gap of 100 bp
* 17868 21613: contig of 3746 bp in length
* 21614 21713: gap of 100 bp
* 21714 25648: contig of 3935 bp in length
* 25649 25748: gap of 100 bp
* 25749 30143: contig of 4395 bp in length
* 30144 30243: gap of 100 bp
* 30244 35339: contig of 5296 bp in length
* 35340 35639: gap of 100 bp
* 35640 39963: contig of 4324 bp in length
* 39964 40063: gap of 100 bp
* 40064 45751: contig of 5688 bp in length
* 45752 45851: gap of 100 bp
* 45852 52881: contig of 7030 bp in length
* 52882 52981: gap of 100 bp
* 52982 61044: contig of 8063 bp in length
* 61045 61144: gap of 100 bp
* 61145 69139: contig of 7995 bp in length
* 69140 69239: gap of 100 bp
* 69240 76404: contig of 7165 bp in length
* 76405 76504: gap of 100 bp
* 76505 83553: contig of 7049 bp in length
* 83554 83653: gap of 100 bp
* 83654 91363: gap of 100 bp
* 91364 91463: gap of 100 bp
* 91464 100416: contig of 8953 bp in length
* 100417 100516: gap of 100 bp
* 100517 108752: contig of 8236 bp in length
* 108753 108852: gap of 100 bp
* 108853 118890: contig of 10038 bp in length
* 118891 118990: gap of 100 bp
* 118991 131762: contig of 12772 bp in length
* 131763 131862: gap of 100 bp
* 131863 143447: contig of 11585 bp in length
* 143448 143547: gap of 100 bp
* 143548 160191: contig of 16644 bp in length
* 160192 160291: gap of 100 bp
* 160292 180075: contig of 19784 bp in length
* 180076 180175: gap of 100 bp

```


* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 176380: contig of 176380 bp in length
 * 176381 176480: gap of 100 bp
 * 176481 190539: contig of 14059 bp in length.

FEATURES

Source
 1. 190539
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-685B14"
 /clone_lib="RP11 Human Male BAC"
 1. 176380
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 176481. 190539
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"

misc_feature
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 176481. 190539
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"

BASE COUNT 61261 a 37428 c 36378 g 55372 t 100 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 190539;
 Best Local Similarity 87.5%; Pred. No. 75;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtgcgaagtcacagaattattt 25
 111111111111111111111111
 Db 111817 TTGTGCAAGTCACAAATTACTT 111840

RESULT 27

AC007564/c

LOCUS

DEFINITION

AC007564 194058 bp DNA PRI 03-JUL-1999

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

AC007564

JOURNAL

Submitted (03-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 1, 1999 this sequence version replaced gi:5263308.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features Listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3369-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality_info/genbank_annotation.html.

QUALSTAT-REPORT-----

----- Summary Statistics -----
 Contig Length: 122887
 Phrap values in estimate: 121646
 Average error rate (BCM-Phrap estimate): 0.000150138
 Fraction of Phrap values less than 40 : 0.0216119
 Number of consensus changing edits: 45
 Number of N's in consensus : 0

----- Consensus changing edits -----
 Position Original+Context Edited+Context
 66363 acatattgt(n)tgntatata acatttgt(n)tgntatata
 66366 tatgtgtg(n)gtacataat tatgtgtg(n)gtacataat
 66426 tatcacaca(n)acacacac tatcacaca(n)acacacac
 71332 cgtcacact(n)taaaagcgt cgtcacact(n)taaaagcgt
 71375 ttcaactgt(n)taaaagcgt ttcaactgt(n)taaaagcgt
 71414 ggtataagat(n)acaaaacat ggtataagat(n)acaaaacat
 78081 tgacttttt(n)ttcttcagc tgacttttt(n)ttcttcagc
 78101 cactttaag(n)ngncaamca cactttaag(n)ngncaamca
 78102 actttaagn(n)gncaamcat actttaagn(n)gncaamcat
 78104 tttaagng(n)caamcatat tttaagng(n)caamcatat
 78108 aagngncaa(n)ncatatac aagngncaa(n)ncatatac
 78109 agnngncaa(n)ncatatac agnngncaa(n)ncatatac
 78119 gttattacac(n)tggtccaa gttattacac(n)tggtccaa
 78194 gttttttc(n)ttcgttgat gttttttc(n)ttcgttgat
 78224 ttcttcacat(n)ttgatttc ttcttcacat(n)ttgatttc
 78272 gttgtttct(n)ttatattc gttgtttct(n)ttatattc
 78989 tattttgt(n)tttgtttag tattttgt(n)tttgtttag
 102112 acatgttgaa(c)ccgtctcac acatgttgaa(a)ccgtctcac


```

* 91237 97282: contig of 6046 bp in length
* 97383 97382: gap of unknown length
* 97383 104072: contig of 6690 bp in length
* 104073 104172: gap of unknown length
* 104173 111143: contig of 6971 bp in length
* 111144 111243: gap of unknown length
* 111244 118031: contig of 6788 bp in length
* 118032 118131: gap of unknown length
* 118132 122562: contig of 4431 bp in length
* 122563 122662: gap of unknown length
* 128512 128512: contig of 5850 bp in length
* 128513 128612: gap of unknown length
* 128613 135250: contig of 6638 bp in length
* 135251 135350: gap of unknown length
* 135351 144540: contig of 9190 bp in length
* 144541 144641: gap of unknown length
* 144641 153087: contig of 8447 bp in length.

```

FEATURES

```

source
    1. .153087
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone="RP11-95F16"
    1. .1284
        /note="assembly_name:Contig9"
    misc_feature
        1385. .3256
            /note="assembly_name:Contig10"
    misc_feature
        3357. .5161
            /note="assembly_name:Contig11"
    misc_feature
        5262. 7722
            /note="assembly_name:Contig12"
    misc_feature
        7823. .9807
            /note="assembly_name:Contig13"
    misc_feature
        9908. .11659
            /note="assembly_name:Contig14"
    misc_feature
        11760. .14961
            /note="assembly_name:Contig15"
    misc_feature
        15062. 17341
            /note="assembly_name:Contig16"
    misc_feature
        17442. .20613
            /note="assembly_name:Contig17"
    misc_feature
        20714. .24483
            /note="assembly_name:Contig18"
    misc_feature
        24584. .27629
            /note="assembly_name:Contig19"
    misc_feature
        27730. .29986
            /note="assembly_name:Contig20"
    misc_feature
        30087. .32919
            /note="assembly_name:Contig21"
    misc_feature
        33020. .35697
            /note="assembly_name:Contig22"
    misc_feature
        35798. .38277
            /note="assembly_name:Contig23"
        clone_end:"7"
        vector_side:"left"
    misc_feature
        38378. .41653
            /note="assembly_name:Contig24"
    misc_feature
        41754. .44674
            /note="assembly_name:Contig25"
    misc_feature
        44775. .49609
            /note="assembly_name:Contig26"
    misc_feature
        49710. .53789
            /note="assembly_name:Contig27"
    misc_feature
        53890. .57453
            /note="assembly_name:Contig28"
    misc_feature
        57554. .61602
            /note="assembly_name:Contig29"
    misc_feature
        61703. .65138
            /note="assembly_name:Contig30"
    misc_feature
        65239. .70266
            /note="assembly_name:Contig31"
    misc_feature
        70367. 75388
            /note="assembly_name:Contig32"

```

```

misc_feature 75489. .80195
              /note="assembly_name:Contig33"
misc_feature 80296. 85960
              /note="assembly_name:Contig34"
misc_feature 86061. .91136
              /note="assembly_name:Contig35"
misc_feature 91237. .97282
              /note="assembly_name:Contig36"
misc_feature 97383. 104072
              /note="assembly_name:Contig37"
misc_feature 104173. .111143
              /note="assembly_name:Contig38"
misc_feature 111244. .118031
              /note="assembly_name:Contig39"
misc_feature 118132. .122562
              /note="assembly_name:Contig40"
misc_feature 122663. .128512
              /note="assembly_name:Contig41"
misc_feature 128613. .135250
              /note="assembly_name:Contig42"
misc_feature 135351. .144540
              /note="assembly_name:Contig43"
misc_feature 144641. .153087

```

Query Match 76.8%; Score 19.2; DB 2; Length 153087;
 Best Local Similarity 87.5%; Pred. No. 74;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgaagtcacgaattatt 25
 Db 116175 TTCTCACAAGTCACGATTCATT 116152

```

RESULT 23
CNS057C0
LOCUS      CNS057C0 158214 bp DNA PRI 11-MAY-2001
DEFINITION Human chromosome 14 DNA sequence BAC R-15E14 of library RPI1-11
ACCESSION AL352981
VERSION    AL352981.4 GI:14041152
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 158214)
            Bellig,R., Petzl,J.L., Vico,V., Dasilva,C., Robert,C., Mincker,P.,
            Brothier,P., Catolico,L., Barbe,V., Pelletier,E., Arliguenaue,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
            Sequencing of the human chromosome 14
            Unpublished
            2 (bases 1 to 158214)
            Genoscope.
            Direct Submission
            Submitted (11-MAY-2001) Genoscope - Centre National de Sequenage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            On May 15, 2001 this sequence version replaced gi:10241556.
TITLE      Genome Center
JOURNAL    Center: Genoscope / Centre National de Sequenage
REFERENCE  Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr

```

The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-74H1
 Downstream BAC (overlapping the SP6 end) : R-6101 (AC=AL163932)

----- Summary Statistics -----
 Assembly program: Phrap; version 2.0
 Quality coverage: 7.27x in Q20 bases; sum-of-contigs

repeat_region 65224..65456
/note="L1ME3 repeat: matches 910..691 of consensus"
repeat_region 65303..65456
/note="L1ME3 repeat: matches 839..691 of consensus"
repeat_region 66620..66721
/note="MER20 repeat: matches 23..125 of consensus"
repeat_region 66973..67866
/note="L1PA15 repeat: matches 897..1 of consensus"
repeat_region 67723..72165
/note="L1 repeat: matches 5390..987 of consensus"
repeat_region 73816..74035
/note="MIR repeat: matches 249..18 of consensus"
repeat_region 74040..74096
/note="L1MCI repeat: matches 1079..1022 of consensus"
repeat_region 74155..74188
/note="17 copies of 2 mer 82 % conserved"
repeat_region 74286..74681
/note="MULTI1 repeat: matches 362..2 of consensus"
repeat_region 76306..76443
/note="MIR2 repeat: matches 138..1 of consensus"
repeat_region 77147..77434
/note="Aluio repeat: matches 16..302 of consensus;
incomplete repeat"
repeat_region 78277..78947
/note="I1 repeat: matches 3606..2975 of consensus"
repeat_region 79099..79390
/note="Alusx repeat: matches 302..1 of consensus"
repeat_region 79413..79776
/note="I1 repeat: matches 2512..2123 of consensus"
repeat_region 79882..80048
Query Match 76.8%; Score 19.2; DB 9; Length 132981;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 ttgtcgaagtcacagaattatt 25
Db 84233 TTGTCAATGACAGATTATT 84256

RESULT 22
AC020546/c 153087 bp DNA HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 14 clone RP11-95F16, WORKING DRAFT
DEFINITION
SEQUENCE, 36 unordered pieces.
AC020546
AC020546.2 GI:9502458
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 153087)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL unpublished
REFERENCE 2 (bases 1 to 153087)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 27, 2000 this sequence version replaced gi:6665582.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0095F16
----- Summary Statistics -----
Sequencing vector: M13; 93%
Sequencing vector: plasmid; 7%

Chemistry: Dye-primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 133287 bases at least Q40
Consensus quality: 139089 bases at least Q30
Consensus quality: 142441 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 149587; sum-of-contigs
Quality coverage: 3.21 in Q20 bases; sum-of-contigs
Quality coverage: 3.10 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1284: contig of 1284 bp in length
1384: gap of unknown length
1285 1384: gap of unknown length
1385 1385: contig of 1872 bp in length
1387 3356: gap of unknown length
3357 5161: contig of 1805 bp in length
5162 5262: gap of unknown length
5262 7723: contig of 2461 bp in length
7723 7823: gap of unknown length
7823 9807: contig of 1985 bp in length
9807 9808: gap of unknown length
9808 11659: gap of 1752 bp in length
11659 11759: gap of unknown length
11759 14961: contig of 3202 bp in length
14961 15061: gap of unknown length
15061 17341: contig of 2280 bp in length
17341 17441: gap of unknown length
17441 20613: contig of 3172 bp in length
20613 20713: gap of unknown length
20713 20714: contig of 3770 bp in length
20714 24483: contig of 3770 bp in length
24483 24584: gap of unknown length
24584 27629: contig of 3046 bp in length
27629 27729: gap of unknown length
27729 29986: contig of 2257 bp in length
29986 30086: gap of unknown length
30086 32919: contig of 2833 bp in length
32919 33019: gap of unknown length
33019 35697: contig of 2678 bp in length
35697 35797: gap of unknown length
35797 35798: contig of 2480 bp in length
35798 38277: gap of unknown length
38277 38377: gap of unknown length
38377 41653: contig of 3276 bp in length
41653 41753: gap of unknown length
41753 44674: contig of 2921 bp in length
44674 44774: gap of unknown length
44774 49609: contig of 4835 bp in length
49609 49709: gap of unknown length
49709 49710: gap of unknown length
49710 53789: contig of 4080 bp in length
53789 53889: gap of unknown length
53889 57453: contig of 3564 bp in length
57453 57554: gap of unknown length
57554 61602: contig of 4049 bp in length
61602 61702: gap of unknown length
61702 61703: gap of unknown length
61703 65138: contig of 3436 bp in length
65138 65238: gap of unknown length
65238 70266: contig of 5028 bp in length
70266 70366: gap of unknown length
70366 75388: contig of 5022 bp in length
75388 75488: gap of unknown length
75488 80195: contig of 4707 bp in length
80195 80295: gap of unknown length
80295 85960: contig of 5665 bp in length
85960 86061: gap of unknown length
86061 91135: contig of 5076 bp in length
91135 91235: gap of unknown length

corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre chromosome X mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChX/>

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 82J11 is at 1 in this sequence. The true left end of clone U134E6 is at 62897.

The true right end of clone U134E6 is at 102829.

The true right end of clone 82J11 is at 132981.

82J11 is from the library RPCII constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.

For further details see <http://bacpac.med.buffalo.edu/U134E6> is from the Lawrence Livermore National Laboratory flow-sorted X chromosome cosmid library LL0XNC01.

FEATURES

source

1. 132981

/organism="Homo sapiens"

/db_xref="taxon:9606"

source

1. 92644

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="RP1-82J11"

/clone_lib="RPC1-1"

repeat_region

/note="LTR repeat: matches 561. 383 of consensus"

repeat_region

/note="11 copies of 2 mer 100 % conserved"

repeat_region

/note="MER3 repeat: matches 208. 1 of consensus"

repeat_region

/note="LI repeat: matches 5281. 4409 of consensus"

repeat_region

/note="LI repeat: matches 34. 893 of consensus"

repeat_region

/note="LIPAZ repeat: matches 34. 893 of consensus"

repeat_region

/note="21 copies of 2 mer 93 % conserved"

repeat_region

/note="LIM4 repeat: matches 1039. 66 of consensus"

repeat_region

/note="LIPAZ repeat: matches 890. 825 of consensus"

repeat_region

/note="LIPAZ repeat: matches 5299. 1333 of consensus"

repeat_region

/note="MER25 repeat: matches 2132. 1556 of consensus"

repeat_region

/note="LIM4 repeat: matches 931. 333 of consensus"

repeat_region

/note="MIR repeat: matches 4. 254 of consensus"

repeat_region

/note="Alub repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

repeat_region

/note="LIPAZ repeat: matches 302. 134 of consensus"

/protein_id="CA06091.1"

/db_xref="GI:2780173"

/translation="PEESPKEPSEVNVNPLVSPACKPLIMHEKEFTSEICGSLW
GVNLISLSTRMLYLMDSKGRADITKLRPPFOIOVLEPLNLLITISGKNRLRYH
LTLWRNKLINNDPEESKROEMLKTEACAKAIKLPFOHEFTYIAIALSKSYHA
WAPKSPESSTAIKVEPTLDKPYVDLAISERLKFESSACGYLIDAESVMDY
TLPNNTLILIDPCIGTGMILFVNAEALSVANQDLKILIMKRDIPSSIAFECTORT
TGWCKALEPSSDRVLESELKRLKILPLCTRODK"

26070. 26296

/note="Alu repeat: matches 51. 279 of consensus"

incomplete repeat"

29261. 29598

/note="TMRK repeat: matches 371. 36 of consensus"

30655. 30695

/note="5S repeat: matches 41. 1 of consensus"

32254. 33127

/note="LIM8 repeat: matches 920. 1 of consensus"

33130. 33431

/note="AluX repeat: matches 296. 1 of consensus"

33435. 34221

/note="LI repeat: matches 5248. 4454 of consensus"

34694. 34898

/note="Alu repeat: matches 301. 100 of consensus"

35592. 35734

/note="MER5A repeat: matches 189. 43 of consensus"

37678. 37969

/note="Alu repeat: matches 10. 301 of consensus"

41035. 41084

/note="25 copies of 2 mer 100 % conserved"

41321. 41354

/note="17 copies of 2 mer 91 % conserved"

45370. 45533

/note="LIM8 repeat: matches 862. 1031 of consensus"

46370. 49114

/note="match: ESTs N30117 M60729 N33939 R98572 R28658
R70758 R62577 H47233 H47234 R62576 R06699 R70808 R28510
H06749 AA331981 AA335043 AA249339 AA247203 H03220
AA331981"

Complement(47194. 47616)

/note="match STRS G28348 G26027"

Complement(49515. 49722)

/note="match STRS AF020167"

50129. 50131

/note="LI repeat: matches 7. 388 of consensus"

50573. 55498

/note="LI repeat: matches 479. 5390 of consensus"

55353. 56235

/note="LIPAZ repeat: matches 1. 891 of consensus"

57119. 57154

/note="18 copies of 2 mer 81 % conserved"

58395. 58799

/note="MLTAL repeat: matches 362. 4 of consensus"

59693. 59994

/note="AluX repeat: matches 302. 1 of consensus"

59837. 59917

/note="MST repeat: matches 48. 1 of consensus"

60007. 60053

/note="MST-INTERNAL repeat: matches 1621. 628 of
consensus"

61117. 61960

/note="LI repeat: matches 4530. 5388 of consensus"

61829. 62702

/note="LIPAZ repeat: matches 1. 887 of consensus"

62721. 62912

/note="MST-INTERNAL repeat: matches 620. 433 of consensus"

63677. 63679

/note="clone U134E6; GGA in this entry; insertion"

64668. 64889

/note="LIM3A repeat: matches 349. 565 of consensus"

FEATURES	Location/Qualifiers
source	1. .155633

Query Match	80.8%;	Score 20.2;	DB 2;	length 155633;
Best Local Similarity	88.0%;	Pred. No. 24;		
Matches 22; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

ttctgcacatcacacagattattc 25
||||| ||||| ||| ||||| ||
TGTGTCACAGTAACATAATTATT 4892

AC020741 191804 bp DNA PRI 09-MAY-2001
Homo sapiens BAC clone RP11-798L4 from 4, complete sequenced.
AC020741
AC020741.4 GI:10864254
HMG.

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 191804)
Suiston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792

2 (bases 1 to 191804)
Marquis-Homeyer,C., Drone,K., Gregory,S. and Boyer,E.
The sequence of Homo sapiens BAC clone RP11-798L4
Unpublished
3 (bases 1 to 191804)
Waterston,R.H.
Direct Submission
Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 191804)
Waterston,R.H.
Direct Submission
Submitted (18-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 191804)
Waterston,R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 18, 2000 this sequence version replaced gi:7631118.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
Summary Statistics
Center project name: H_NH0798L04

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPheron, Department of Genetics, Washington University, St. Louis,
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPc1-11 human BAC library was made from the blood of one male

QY 1 atctgcacatcacagaattatt 25
 |||||
 DB 4052 ATTGTGCAACTACATGATTATT 4076

RESULT 17
 AF250770 965 bp DNA BCT 31-JAN-2001
 LOCUS
 DEFINITION Uncultured bacterium pCosHE1 DAPA-aminotransferase (btoa) and
 biotin synthase (biot) genes, partial cds.
 ACCESSION AF250770
 VERSION AF250770.1 GI:12620104
 KEYWORDS
 ORGANISM
 SOURCE uncultured bacterium pCosHE1.
 Bacteria; environmental samples.
 REFERENCE
 AUTHORS 1 (bases 1 to 965)
 TITLE Fitcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streitt, W.R.
 JOURNAL Direct cloning from enrichment cultures, a reliable strategy for
 MDLINE isolation of complete operons and genes from microbial consortia
 20575196 Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
 PUBMED 11133432
 REFERENCE 2 (bases 1 to 965)
 AUTHORS Entcheva, P., Liebl, W. and Streitt, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1. 965
 /organism="uncultured bacterium pCosHE1"
 /db_xref="taxon:143796"
 /clone="pCosHE1"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(<1. 438)
 /gene="btoa"
 /gene="btoa"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase; similar to btoa"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase"
 /protein_id="AAG60563.1"
 /db_xref="GI:12620105"
 /translation="MTODLAFDDQIHMPYTSMTREXPVYPVASAXHCOORRRIAD
 GMSWMAIHGYNHPRILNAKAOIEQSHVDFGITHOPAVDLCRRXVANTPDALEC
 VFLADSGVAXEVXDENCASTAPLPADRPERRHXAFIFXLO"
 525. >965
 /gene="btoa"
 525. >965
 /gene="btoa"
 /note="similar to btoa"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AAG60564.1"
 /db_xref="GI:12620106"
 /translation="MXHHARWTMSQVTELEFKPFLLEMFEAQVHROHPDRIYVOST
 LSIKTGACPEDCKYCPQSAVKYTGLESERLMEYQVLDASRKKNAGSTRFCGAAW
 KKPHIVCPLEDMVGVKMGCEALYDPRHARRRXXSACLLA"
 BASE COUNT 230 a 256 c 285 g 180 t 14 others
 ORIGIN

Query Match 81.6%; Score 20.4; DB 1; length 965;
 Best Local Similarity 95.5%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 tctgcacatcacagaattatt 24
 |||||
 DB 550 TGTGCAAGTCACGATTATT 571

RESULT 18
 AC025510
 LOCUS
 DEFINITION Homo sapiens chromosome 4 clone RP11-673L23 map 4, WORKING DRAFT
 SEQUENCE, 13 unordered pieces.
 AC025510
 VERSION AC025510.3 GI:8077120
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 155633)
 TITLE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 155633)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, B., Bouckgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodgson, S., Domini, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glend, S., Goyette, M., Fitzhugh, W., Gage, D.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lebecky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Melidum, J., Menus, L., Milova, T., Miranda, C., Mieng, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Tigillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 330 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced by: 7331573.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L8220
 Center clone name: 673.L23
 Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 149314 bases at least Q40
 Consensus quality: 152581 bases at least Q30
 Consensus quality: 153851 bases at least Q20
 Insert size: 157000; agarose-fp
 Insert size: 154433; sum-of-contigs
 Quality coverage: 4.8 in Q20 bases; agarose-fp
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 13 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence

```

CDS
/gene="putu"
<1..502
/feature="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"
/db_xref="GI:12407618"
/translation="PERRISTFOGLIPARICWGLEWROKLGIAFNEMVRRGEVAPIV
GSDHLDSCVSPSPNRETEMDGSDVSDWPLNLINFGATVNSLHGGCGMCP
SOHAGMIVICDGTDDAARILARVILHNDPATGVMRHADAGYDIALDCAEGGLNPLMIA
ATOGQH"
502..2031
/feature="putu"
502..2031
/feature="putu"
/feature="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAC53586.1"
/db_xref="GI:12407611"
/translation="MNALITPGLTLTAOLROVWQPIOLTI DESAHEAINDSVACV
AIVAEGRVYGINTEGFLLAOTRIATHLENIQSRSLVSHAAGVEPLDDIVLMMV
LKINSLARFSGIRLSVIOALIAVNAAGVYSDPKGSYGAGDLAPLAHMSLTLLE
GKARYGEMLPATALOKAGLAPVTLAAKEGLALINGQASTAFALRGLEAEDLEAS
AVCGALTEAVIGSRPRPDARLHEVROGGOIDAAALFRHVLTDISAIESHNCDK
VDDPYSLRCOPQVMACTIOMROVAEVLIVESNAVDNPLVFAENEMVRCNFAEP
VAMADNLALATAETGALSERIALMCKHMSLOPELVKNGVNSGEMIAOVTAAL
ASENKLCHPHHTSYDPCPSKANDSDHSMAPACGRRLREAGNTRGLAVENLAAC
OGADLRDGLTSSPLEEQAROSCEGYAHYDDDFPAPDIEAAILSLNKSLVGLLPF
L"
complement(2130..2606)
/feature="OprL: similar to hypothetical 17.1 kDa protein in
modc-bioa intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC53587.1"
/db_xref="GI:12407612"
/translation="MKLISQDLRDGDKLPLRHVNGMGYEGDNI SPHLAMDEYSGIK
SEVYTCYDPAPTPDSSGMWHITVANIADPRVLPDGGSSSLVALPEGAVETRIDPGAG
YGAAPKGETHRTIFTVHAIINVERIEVDENAGAVGCRVHHTTIGSATIMYS"
complement(2651..3940)
/feature="bioa"
complement(2651..3940)
/feature="bioa"
/feature="7,8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase Bioa"
/protein_id="AAC53588.1"
/db_xref="GI:12407613"
/translation="MTTDDLAFPDRHIMHPTSTOPLPVYPVVEAHGCELTIASGEQ
LYGMSWMAIHGYNHPRILNAAKAOIDRSHVMGGITHOPAVDRCITGAMSVCDPNS
LECYFPADSGSAVEVAMKALDYMAHAKGSSRQRTITPNCYHCDTREGAMSVCDPNS
MHSILMKYLPENLPAPOSRDEGAMDEMOWGLAPLMAHRRHETAAVITLTPVQAG
GRMYHPEMLKRIIRKMDCREGLILADELATGCRGKLFACCHAGITPDLTGKAL
TGITMTSATLITRQVAETISNGEAGCFMGPMPGLACAAVASELALLESGEWD
OGATIESOLOBELVARDSEFVADVPLGAIGVETTHPVMMAALORFEFVARGVWVP
FGKLIYMPVYIRPDLSKLTAVVAVOAPAHFTR"
4029..5069
/feature="bioB"
4029..5069
/feature="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAC53589.1"
/db_xref="GI:12407614"
/translation="MAHQARVTLISQVAFLEFKPLLELFEAQOITRHOHPDQOIVST
LTSIKTACPEDCKYCPQSAARYKTGLESRIMEVEGYLESARAKNAGSTRFCMAAW

```

```

gene
/feature="bioC"
/feature="7-keto-8-amino-pelargonic acid synthetase: similar
to BioC"
/codon_start=1
/transl_table=11
/product="7-kAPA synthetase"
/protein_id="AAC53590.1"
/db_xref="GI:12407615"
/translation="MTWQQRIDDAALDARRAADALRVRTPYENGAGRFLIREDFRCNF
SGNDYLGISOHPEIRAMQOAGAEQYSGSGSHVGTTHAQALEHLAGMVGYPRA
LITSGFRAAOALIAALTDEKEDRIVADRILSHALISAHSPLAOLRRFAHNDVAODS
LDRKCHGQDLVTTEGIFGSMDSAPGLGVTHAAOBSGSLVDDAHGIGVTPREGG
SAAQGVGPELVVTEFGVGAALVGSADAVADITLOPARLITSTSPRAQAVL
LAAPNVRSDENARQRLAALIDQFPAGVRLPVQITASOSAIDPLVGENAKALHL
AQRLEKGIWVTAIRPTVPAGTARLRLTLAHOADITLALVALHAGE"
6207..6962
/feature="bioC"
6207..6962
/feature="bioC"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis bioC-like protein"
/protein_id="AAC53591.1"
/db_xref="GI:12407616"
/translation="MSOVKQALIAAFGRGAOSRRDALOROSADGLRALLENFPA
SYLDGCGPFGNSIALRQGTGVYTAQMLCPMEFARRNOAOSLOADIEALIPDA
QPSLAWSHLAVOXCASLPRALAEMLRVYTOIGKARFTLLANGSLPELNQMOAVDOP
HANRPLSGOQVNDALSGWQYCAVQVTLTFDDVLSMHSLSKLGATHLHAGRAKPL
TRGOLQOLQAQAMPQOAGKFLSYOLFGLIERD"
6940..7680
/feature="bioD"
6940..7680
/feature="bioD"
/codon_start=1
/transl_table=11
/product="gethiobiotin synthase bioD"
/protein_id="AAC53592.1"
/db_xref="GI:12407617"
/translation="WELSNVBERYFYVGTOTRENGKTYASTALLOAARALGRHTAGYRP
VASGSEMTAQGLRNSDALALDRNSLSALSTFRAVNPYTPAEPSTPHIYSADBOPIERA
TLISGLRALLETQADWLVLEGAGWETPLISODTRADVMARLPLVILVGYKGC INH
AMLTQOALIRQAGLHFGAGVANGVYVGRRAHEMATLRLVLSAPLIGELIPWLSBOADT
SAVGQYLDLTALTTPSPSSGOSAGHPAR"
complement(7619..8227)
/feature="eisa"
complement(7619..8227)
/feature="eisa"
/feature="putative urea/short-chain amide ABC transporter"
/feature="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-like protein"
/protein_id="AAC53594.1"
/db_xref="GI:12407619"
/translation="DCXILTRXGEXKXLYSCITAGLPIASGTRWHEARGAONQGTET
XVVPQDRRTFESOLVTEENIHTAONAKGJAGSLSDVVEFEPELPLQOLASGLND
OOYOLANALATVTRPRLILDEPSRGTOQLFTVKLAOLLRNRELGLSVLLAEQHS
LIRVADRCLLYRGRNVAOGVSELDPLLAHVMGPASAO"

```

```

BASE COUNT      1512 a      2642 c      2605 g      1453 t      15 others
ORIGIN
Query Match      87.2%; Score 21.8; DB 1; Length 8227;
Best Local Similarity 92.0%; Pred. No. 3.6;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

source
1. .5526
/organism="uncultured bacterium pcoshe2"
/db_xref="taxon:143797"
/clone="pcoshe2"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(52. .528)
/note="Orf1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kda protein in modc-bioa
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDSDKLPRIHVENCMGYDGNISPHLAMDVPAGTK
SFVVCYDDPAPGSGMHVYNNVLPADRVLPDQSGSLVAMPDGLVTRDTRFGKAG
YDCAAPPKSETHYITFYHALDIERIDVDEGASGAWGNVHFSLASASTYMS"
complement(587. .1876)
/gene="bioa"
/complement(587. .1876)
/note="7.8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase Bioa"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFDQRIHNPITSMPSLPDVPVSAESCECILSDGR
LYDGMSSMAALHGNHNPOLNAKMSQIDAMSHVFGGTTAIPALCLKLVAMPQP
LACVLADGSSVAVEAMKALQYQAKGAEORFLTEPENGHGTFGMSVCDDNS
MISLWKGVLPEMLFAPQSRMDGEMDEDMGFARLMAAHNHEIAAVIIEFYOGAG
GARMYHEWIKLRKICDREGILLADLATEFGRTGKLPCEHNEIETLQICKAL
TGGTMTLSATLTTRTREAETISNGEAGCGMHGPTFMGNPLACAAASIALIESGDMW
OYADLEVOLEQALPARADAWADRVICATVYTTTPVMAALQKFFVEGVAWRP
FGKLIYMPPIITLPOOLRLTAANVRAVODETFPCQ"
1963. .3003
/gene="bioa"
1963. .3003
/gene="bioa"
/codon_start=1
/transl_table=11
/product="biotin synthase Bioa"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPMTLSQVTELFEPKPLDLLFEAQOYVHROHDPROYVST
LISITGACPEDEKCYPOSRYKTGLEAEERLMEVEOYLSARKAKAGSTPFCGAAH
KNPNERDMPYLEOMYGKDLGLEACMTIGTSSQORLNAAGLDIYNHLDTPES
YGNITTRTYOERLDITLKVVDAGIKVCSGTVGIGETKDRAGLLQLANLPTPPES
VPINMLVVKSTPLADNDVDAFDFTRITAVARIMPTSVYLSAGREONNEOTQAMC
FMAGANSTFYGCCKLTTPPEDEKDLQLEFKLGINPOTAVLADGNEOQRLQALMT
PDTDEYNAAL"
3000. .4154
/gene="bioa"
3000. .4154
/gene="bioa"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase Bioa"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSMOEKTNALDARRAADALRRYPAQAGRWLVADROYLNF
SSNDYLGISHHPQIIRAMQGAEOFGTSGSGSVASVYHQALEBELAEWLGYSRA
LLETISGRANQAVIAAMAKEDRIADRLSHASLLEASLSPSOLRRFANDVYHLAR
LAPSPGOOLVTEGVFSMDGSAPIAEIOOVQOHNGLMWDAGTGVIGGOGG
SCWLOKVKPELLVTFPGFGVSGAVALIEQOVTYDYLQFARHRIYTSMPPOAOUL
RASLAVIRDEGDARREKIVSLIHPFAGVODLPETLADSCSAIOPLIIVDNSRALDL
AEKLRQGCWCWTALRPPTVAGTARLRLTLTAHEMODIDRLLEVLHGNG"
4141. .4896
/gene="bioa"
4141. .4896
/gene="bioa"
/gene="bloc"
4141. .4896
/gene="bloc"
4141. .4896
/gene="bloc"

gene
1. .8227
/organism="uncultured bacterium pcosas1"
/db_xref="taxon:143793"
/clone="pcosas1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502

FEATURES
Source
1. .8227
/organism="uncultured bacterium pcosas1"
/db_xref="taxon:143793"
/clone="pcosas1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502

RESULT 16
AF248314
LOCUS
DEFINITION
AF248314 8227 bp DNA BCT 24-JAN-2001
uncultured bacterium pcosas1 urocannase-like protein (hutu) gene,
partial cds: histidine ammonia-lyase-like protein (hutu), DAPA
aminotransferase Bioa (bioa), biotin biosynthesis BioC (bioB), 7-KAPA
synthetase (bioC), biotin biosynthesis BioC-like protein (bioC),
and dehydrobiotin synthase BioD (bioD) genes, complete cds: ABC
transporter-like protein (elsa) gene, partial cds; and unknown
gene.
AF248314
AF248314.1 GI:12407610
uncultured bacterium pcosas1.
Bacteria: environmental samples.
1 (bases 1 to 8227)
Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streitt,W.R.
Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL
REFERENCE
PUBMED
2 (bases 1 to 8227)
Entcheva,P., Liebl,W. and Streitt,W.R.
Direct Submission
Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
Location/Qualifiers
1. .8227
/organism="uncultured bacterium pcosas1"
/db_xref="taxon:143793"
/clone="pcosas1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502

```

```

/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNKTVICOGFTSGSGTFHSEDAIATGTMKGAVTPGKG
GTHHGLFVFNTPRRAVAATGATGATVIVAPAFPCDSILEAIDAGIKLITITGIPIT
EDMLTVKRLDEAGYRMTGPNPCVITPTECEGKIGTGPRIHKKPKGVIGVSRSGLLTVE
AAVKOTTDYFGOSTCVGIGDPIPOSNFDITILEMEKDPQTEALVMIGETGSAEEFA
AAYKEHVTKRPVAGYIAGVTAPKGRMGHAGAIAGGKGTADERFAALEAGVTVRS
LADICEALKTVLK"
/2781..3686
/gene="ECS0755"
/2781..3686
/gene="ECS0755"
/translation="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
Escherichia coli g11417043|sp|P32064|CVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKITPKTELLVTFEVVARHESYTKAEELALJOSAPFQVSAL
BEFTKPLFSHRSKREINDNGKYIVKTEINKLEPNTIMIOPTVQVIELAVN
PISTHWLIPNHEFTKLPDITIVNHSLANNQDPLNREDAVIMRENPCAMAEVY
LDEEETLIPVCSGLAMSDOKLSVABLELPLHGSTRTTCMEERFALSSVSSLVN
NGRFDLSMLIAAVRSNLGVALLPFAIQLHDLDSGDWVIPCQVPIRTGNRIKTMQE
EKSDSPHLOQFREMILAKSVPOEM"
/complement(3720..4322)
/gene="ECS0756"
/complement(3720..4322)
/gene="ECS0756"
/translation="probable cob(1)alamin adenosyltransferase, similar
to cob(1)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g1151481sp|P13040|BtUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEERHVOOROKLKEOVDPRAVAATEKKGLIVFTGNGK
KSTAFRTVTRAVGCHKTGCVNOYIKGONDNGYNLPLGVFHHMGTGPFMEQNR
QADIDAKKEVSESKRLADKRYDLVLDLITMLAYHILDTREVIASLQNRPAQDSV
IVRGCHSQILKMDVSEIRPVKHAIFMGIOAQPGJIM"
/complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/gene="ECS0757"
/translation="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g11205981sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFITMOEPFLONKKGTEYLLISDQHTVTELDGEVYKIAPE
AULIARQATYERASFIRSAHLOQVAILDQOASSDQVVALQOLLRMNVEVSKGLP
NCQDTGATVASKQOIVTGNDALSLKGTSTQENNLRSQAPLDMTEVNTQ
TNLPAQIDISAVAGDEYHPLCVNKGGSANKAALYQETKSLQPEKILATLEIKMSL
GTACPPYHIAFVVGGLSADQTLIAKLASTKYNDLPISGNSQOAFRILEKYL
EASQGFIGAOGFGKCPAHDRIVRIPIRHGSCPIAALISDADRINKAKINHGIML
EKLHNPGQYIIPASLRENNAHQVQDLNRPDLVMDLARLPVGTFRVSLSGPIVAVR

```

```

DIAMAKIKARLDSGEPPEYLKHHIIVYVAPKTPENMGSLGFTTGRMDYIDTF
OAGSGSLVMSKDSRQOVADACHKGCFNLISIGGAALLQAEVYKSLRCLLEYELG
MEAVMHEVENLPAFLIVDDKGNFFSOFEOOHRASCAPAH"
/complement(6092..7372)
/gene="ECS0758"
/complement(6092..7372)
/gene="ECS0758"
/translation="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671sp|P24943|GLT_BACG percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKRISLTWITLALVIGMTIGVILNNTASPEAKLYAOEISFT
TIFLRILKMTIAPLVSTLVVGIKMGDAKALGRITSKTLFETCASLSIALGITY
NFWPRTGIVNVAHGAETGVAADEPTLKVFTISHAFPTVDAHANEILOIVSTI
FLGSLTAIGCKSAIVHALDSLAAHMLKLTGYMLFAPLTPVAISALIAERGLAVM
VSAGIFMGEFEFTMLLVLTIGLAIYVGPICIRLTPRALSPALLAFTTSSSEAF
GTLEKLEPQFVSPKIASFVLPIGVSNLVGSMVCSFATVFIQACNINHSISEQITM
LTIIMLSKMGAGPRASMYIAATINQFNPAGLILMGVDPFLDMGRSATNVMSN
AMGANVSRMGEGHEGECGKALKPMSVNL"
/complement(7533..7850)
/gene="ECS0759"
/complement(7533..7850)
/gene="ECS0759"
/translation="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.099;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 attgtcgcaagtcacagaattatt 25
DB 92334 ATTGTGCAAGTCACAGATTATT 92358

```

RESULT 15
AF250776
LOCUS
DEFINITION
AF250776 5526 bp DNA BCT 31-JAN-2001
Uncultured bacterium pcosHE2
modc-bioa intergenic region, DAPA-aminotransferase BioA (bioA),
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
dethiobiotin synthetase BioD (bioD) genes, complete cds, and
dethiobiotin synthetase BioD (bioD) gene, partial cds.
AF250776
AF250776.1 GI:12620124
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
uncultured bacterium pcosHE2.
Bacteria; environmental samples.
REFERENCE
1 (bases 1 to 5526)
Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streifl,W.R.
Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE
PUBMED 11133432
2 (bases 1 to 5526)
Entcheva,P., Liebl,W. and Streifl,W.R.
Direct Submission
Submitted (31-MAR-2000) Microbiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
Location/Qualifiers

KGNRTYRVVATDNKPCVTRSHSGCFNLHLRLGMPKGEISRIEMLNAPSAA
 ERGIVHSNRPDLPPVAPPELPSVDYKNSVGNVYIGKGAUVVEDADATKYIKM
 FTTSSNEETSEVCEPNOYCGSAEAKIYGNNGDIIGIMKINCESLINSISLP
 AHAHYDMEDRLKGLILEVDTTEITVNLIDRANKENPIDISSYNSDRKSBSQIMQ
 SYHGCKODLSIVLSKTI"

gene
 /gene="20990"
 6462..7160
 /gene="20990"
 /function="orf; Other or unknown (Phage or Prophage Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AG55143.1"
 /db_xref="GI:12513758"

/translation="MRPTSLNVLVHSSSSSSSDTIESLVKASSVOMIKNNPOLRF
 QGDHNIYQIEALDK IGSTETRYLNAIEISIKLSETVIHLNSSLGVMAHD
 IAEINHRGTSDFHCNLAIVEYPCGEGISVDPHATVPHIELHFNHNGERLAVES
 SRAESQKSPLLLEAFRTVGLGAFSEEVLSKPKHEIGMPRTSPYKXSALIHDDNT
 VSLGFOOVLRLHPL"

gene
 complement(7668..8144)
 /gene="yphb"
 /note="20992"
 complement(7668..8144)
 /gene="yphb"
 /function="orf; Unknown function"
 /note="Residues 1 to 158 of 158 are 99.36 pct identical to
 residues 1 to 158 of 158 from Escherichia coli K-12 Strain
 MG1655: B0773"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AG55144.1"
 /db_xref="GI:12513759"

/translation="MKLISNDLRDCKLPHRHVENGMDGSDNISPILAMDVPAGTK
 STVATCYDADPTGSGMHWVNLPAOTRYLPQCGSLVAMPDGVLOTYTDGKTG
 YDGAAPKGETHRYETVYHALDVERIDVDEGASGAMGVFNHSLASASITAMS"
 complement(8203..9492)
 /gene="bioa"
 /note="20993"
 complement(8203..9492)

CDS
 Query Match 100.0%; Score 25; DB 1; Length 13501;
 Best Local Similarity 100.0%; Pred. No. 0.092;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacgaattattc 25
 |||
 Db 9602 attgtcgcaagtcacgaattattt 9626

RESULT 14
 AP002553 297816 bp DNA BCT 07-MAR-2001
 LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
 ACCESSION AP002553 BA000007
 VERSION AP002553.1 GI:13360211
 KEYWORDS
 SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
 DNA.
 ORGANISM Escherichia coli O157:H7
 Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 (sites)
 Makioka, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
 Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
 Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
 Sasakawa, C. and Shinagawa, H.
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak

JOURNAL
 MEDLINE
 20198780
 REFERENCES
 AUTHORS

2 (sites)
 Ohnishi, M., Murata, T., Nakayama, K., Kubara, S., Hattori, M.,
 Kurokawa, K., Yasunaga, T., Yokoyama, K., Makioka, K., Shinagawa, H. and
 Hayashi, T.
 Comparative analysis of the whole set of rRNA operons between an
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
 Escherichia coli K-12 strain MG1655
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
 20557356

3 (sites)
 Yokoyama, K., Makioka, K., Kubota, Y., Watanabe, M., Kimura, S.,
 Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
 Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
 Shinagawa, H.
 Complete nucleotide sequence of the prophage VT1-Sakai carrying the
 Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
 O157:H7 strain derived from the Sakai outbreak
 Gene 258 (1-2), 127-139 (2000)
 20564182

4 (sites)
 Hayashi, T., Makioka, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
 Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
 Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
 Ogasawara, N., Yasunaga, T., Kubara, S., Shiba, T., Hattori, M. and
 Shinagawa, H.
 Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12
 DNA Res. 8 (1), 11-22 (2001)
 21156231

5 (bases 1 to 297816)
 Ohnishi, M., Kurokawa, K., Makioka, K., Yasunaga, T., Shinagawa, H. and
 Hayashi, T.
 Direct Submission
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
 Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
 Japan (E-mail: kengen-info.osaka-u.ac.jp,
 URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
 Fax: 81-6-6879-2047)
 genome project.

COMMENT
 FEATURES
 LOCATION/Qualifiers
 1..297816
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:8334"
 79..1245
 /gene="Ecs0753"
 79..1245
 /gene="Ecs0753"
 /note="Similar to SUCC_ECOLI g111786948 percent identity
 100 in 388 aa (conserved in E. coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not experimental
 /product="succinyl-CoA synthetase beta subunit"
 /protein_id="BAB34176.1"
 /db_xref="GI:13360212"

/translation="MNLHEYQAKOLFARYGLPAPVGYACTPPREAEENAKIGAPVW
 VKQVHAGRGKAGKGVNYSKEDIKAEFENMLGRIVYOTDANGOPNOITVEAT
 DIAKELYLGAVIDRSRRVPMASFGGEIVAEPEHLHKVALDPIGPMYOG
 RELAFKLEGGKLVQOFTKIFMGATITFLERDLALIEINPVTITKQSGILCDKGLGA
 DGNALFPRODIREMDROSGDEPREAQMOMELNYALDNGICGVNGLAGTMDIV
 KLHGEPANFLDVGSGATKREYTERFAKIIISDQKAVLVNIFGIVRCDLIADDIIG
 AVALGVNVPVAVRLEGNAELGAKLADSGINTIILAGLIDAAQOYVAAVEGK"
 1245..2114
 /gene="Ecs0754"
 1245..2114
 /gene="Ecs0754"
 /note="similar to SUCC_ECOLI g111786949 percent identity
 100 in 289 aa (conserved in E. coli K-12)"
 /codon_start=1

CDS complement(147..1430)
 /gene="ydhc"
 /function="orf: Not classified"
 /note="f427: 98 pct identical to fragment YBHC_ECOLI
 SW:P46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative pectinesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVRLALALAFGVTLLACSSTPPDQPSDQTAPGTSRPI
 LSKAEQNTDAQHFASLTTPGAAMPSPITLPAQDFVVGPAQTQVTHITQAVD
 AAIKRTNKROYIAVMPGEYOGTVVPAAPGKITLIGTEKPIDVIGISLDGSKPA
 DMKHDVNPGRKYPGKPAWYVSCOSKRSISIGVCSAVESWNNGLQLOLTIET
 LQSDVAGNHPVALRTDQDOYQINNVNITIGRONTFPVNSGONRLTNRPLT
 NSYIEGDVDIVSGRGAVVDFNTEFVNSRTOCPAFAVTPAPLSTNIYTGFLAVNRFN
 AFDGVAOIGRSLDDANNGOVIRDSINIGFNTAKFWADAVISNRPFGNTGSD
 DNDDELRLNLTNTNRYNRMETYNRRGVSKVAEAKK"
 complement(1507..1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 /complement(1582..2058)
 /gene="ydhb"
 /note="b0773"
 complement(1582..2058)
 /gene="ydhb"
 /function="orf: Unknown"
 /note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf: hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDGDLPKRRHVENGMGVDGDNISPHLAWDVPAGTK
 SEVVTCYDPDAPATGSGMHWVYVNLPAATRVLPOGSGLVAMPDGVLOTREDFKIG
 YDGAAPKGETRHYIFVTHALDIERIDVDEGASGMVGFNVHSHASASTTAFPS"
 complement(2108..2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117..3406)
 /gene="b10A"
 /note="b0774"
 complement(2117..3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f429: 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDLAFDORHIMHPYTSMTSPUPYVVSARGCELLISDGR
 LVDNSSWMAIHGYNHNPOLNAAMKSOIDMSHMRGIGTHAPALIELCRKLIANTTPR
 LECEVLADSGSAVEAMKMLQYQAKGAEGRQFLTRNGYHEDTGTGASVCPDSS
 MHSILKGYLPENLFAAPQSRMGDEMRDVGARILMAHREIAVAIVLEPIVOGAG
 GMRWYHPEWLRKIRKICDREGILLIAELIATGERTGLACAEHAEIAPILICGKAL
 TGTMTLSATLTTEVAETISNGEAGCFMHPGMPPLACAAANASLALIESDWOQ
 QVADIEVOLREOLAPARDAEMVADVRLGAIQVETTHPYNMAALQKFVEGQWIRP
 FGKLIYMPYIILPOOLORLTLAAVNRAVODETFPCQ"
 complement(2193..2221)
 /gene="b10A"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411..3450
 /note="central position to b10b promoter: -20"
 /bound_moiety="B10b documented site"
 3411..3450
 /note="central position to predicted promoter: -20;genetic
 evidence for the site"
 /bound_moiety="B10b predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"

protein_bind /bound_moiety="B10b predicted site"
 complement(3411..3450)
 /note="central position to b10a promoter:50"
 /bound_moiety="B10b documented site"
 3413..3441
 /note="factor Sigma70: promoter b10b: documented -1 at
 808525"
 complement(3447..3473)
 /note="factor Sigma70: promoter b10A: documented +1 at
 808515"
 3493..4533
 /gene="b10B"
 /note="b0775"
 3493..4533
 /gene="b10B"
 /EC_number="2.8.1.-"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o346: 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPRWTLISQYTELFEKPLDLLFPAQOYHROHPDPOYVST
 LLSIKGACPEDEKCYCPSSRYKRGLEAEPLMEVEQYLESARAKAGSTRGCGAW
 KNPERMPYLEQWQVCKAMGLEACHTLCTLESQAORLANGLDYNNHNDTSEF
 YGNITTRTOERLDLLEKVDAGIVCSGIVGLEIYKDRAGLLLOLANLPTPEF
 VPINMLVKVKGCTPLADDVDADPFRITTAIVARIMPTSVYRSLASGEQNNBQTOAMC
 FMAGANSIFYGCKLLTTPNPEDEKDLQLEFRKILNPOQVAVLAGDNEQOORLEQALMT
 PDIDVEYNAAL"
 4530..5684
 /gene="b10F"
 /note="b0776"
 4530..5684
 /gene="b10F"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o384: 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSWQEKINAAALDARRAALARRRYPVACAGRWLVADROYLNF
 SSNDYIGLSHHPIIRAWQGAQEQFEGISGSGSHVSGSVYQALEELAEMLVYSRA
 LFIISGFANQAVAIAMMAKEDRIADRLSHSLAASLSPSOLRRFAHNDVTHLAR
 LAFSPGQOMVTEGVSFMSDGSAPLAEIQOVTQOHNQMLVDDAHGTVYGEQGRG
 SCWLQKVPKELLIVTFRGKGFVSGAAVILCSSTVADYLLQFAHRLIYTSMPRAQOAL

Query Match 100.0% Score 25; DB 1; Length 11022;
 Best Local Similarity 100.0%; Pred. No. 0.092;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattatt 25
 |||||
 Db 3516 ATTGTCCGACGATCAGACATTATT 3540

RESULT 13
 AE005258
 LOCUS 13501 bp DNA BCF 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155.
 ACCESSION AE005258 AE005174
 VERSION AE005258.1 GI:12513751
 KEYWORDS
 SOURCE Escherichia coli O157:H7 EDL933.
 ORGANISM Escherichia coli O157:H7 EDL933
 Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 13501)

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaattattt 25
 140 ATTGTGCAAGTCACAGAATTATT 164

RESULT 10
 LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION AR101809
 VERSION AR101809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES
 source
 1. 5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaattattt 25
 140 ATTGTGCAAGTCACAGAATTATT 164

RESULT 11
 LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION AR101810
 VERSION AR101810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES
 source
 1. 5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaattattt 25
 140 ATTGTGCAAGTCACAGAATTATT 164

RESULT 12
 LOCUS AE000180 11022 bp DNA HCF 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 U00096

VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12
 Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B., and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHRP). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 [e-mail: mark@camber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source
 1. 11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1. 130
 /note="RBP (repetitive extragenic palindromic) element;
 contains 4 REP sequences"
 complement(147. 1430)
 /gene="ybhC"
 /note="b0772"

repeat_region
 gene

KNPHERDMPYLQMOVGVKAMGLEACMTLCTLSESOAQLANAGLDVYNNHNDPTSPRE
YGNITTRVYQERLDPLFKVRDAGTICVCGSGTGCETVKKDRAGLLQLANLPTTPES
VPIINMLVKVKGPLANDNDVDADFIIRITAVARIMPTISVRLSAGREMDQOTQAMC
FMAGANSITFGCKLITTPNEEDKDLQFLFRKLGILNPQOAVLAGNEQOORLEOALMT
PDTDEYVNAAL"

RBS
2284. .2297
/standard_name="BIOC RBS"
2295. .3050
/gene="BIOC"
2285. .3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKKQALAAAFGRAAHYRQHADLQROSADALILMPKRYT
HYLDACGCGWMSRHWRERHAQVTAALDSPMLVOARQADADHYLAGDIESLPLATA
TRDLAMSLNAVOMCGNISTALRELYRVVPCVVAFTTLVQSLPELQAOAVDERP
HANRFLPPEIEFOSLGVYHOHIQDITLMPDALSAMRSKLGIGATHLHGRDRIL
TRSOLORLQAMPQOQGRYPLTYHLFLGVIAAE"

RBS
3742. .3752
/standard_name="BIOA RBS"
3750. .5039
/gene="BIOA"
3750. .5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDLAFDQRIHNPPTSMTSPLYPVVYVSAECCELILSDGR
LYDQSSMAATHGTHNPQLNANKSQIDAMSHVMEGITHAPALELCKRLVAMPPOP
LECVLADSGSYAVEAMKMALOYQWQAGARQRLTPFRNGYHGDTPGAMSVCPDPS
NHSILWKGYPENLFAPOSRMDGEMDERMDGAFRLMAARHEIAVITPIYOGAG
GMRVHPEMLKIRKICDREGILLADELAGFGFGTKLFCACHEAETAPDILCICAL
TGGTMTLSATLITREAVETISNGEACGPMHGPPTGNPLACAAANASLILIESGWOQ
OVADILEVOLREGLARABENYADVIRGALGVETTHPVMAALQKRFYBQWVIRP
EKKLILPEPTILLPOQLRLTAANRAVODETFEQ"

RBS
5088. .5100
/standard_name="ORFI RBS"
5098. .5574
/gene="ORFI"
5098. .5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRQDKLPHRHVFNMGSGYDGNISPHLAWDDVPAGTK
SFVYCYDPAAPGSGMHWVYVNI.PADTRVLPOGFGSGLYAMPDGLQTRTDGKIG
YDGAAPKPGETHRITFYVHALDIERIDVDEGASAMGFVNHPSILASASTTAMFS"
5583. .5605
stem_loop
5583. .5644
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 atttgcgaagtcacagaattatt 25
|||||
Db 140 atttgcgaagtcacagaattatt 164
|||||

RESULT 9
A93679
LOCUS A93679 5872 bp DNA
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
AUTHORS 1 (bases 1 to 5872)
TITLE Birch O. and Brass J.
JOURNAL Biotechnological method of producing biotin
Patent: EP 0798384-A 6 01-Oct-1997;
LONZA AG (CH)

FEATURES
source
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154. .2308
/gene="BIOF"
1154. .2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQKINALLDARRAADALRRRYVPAQAGHMLVADROYLNF
SSNDYLGSLHHPOIIRAMQOAGPFGISGSGSVGYSVALLEELAEMLYGSRA
LFTISGPAANOAVIAAMAKEDRIADRLSHSLLEASLSQSLPRFANVYTLAR
LNASPCQGOOVYTEGVSMDSASPLAEIQVYTOQNHMLAWDDAHGTVIGBOGK
SCWLOKYPKELLVYTRGKRGVSGAVALICSSYADYLLDFARHLITSTMPPAOAOL
RASLAVYTRSDGPARREKDALITREKAGVQDLPFLADSCSAIQPLIVGDSRALOL
AEKLRQOGCWVTAIRPPYVAGTARLRLTLTAHEMODIDRLLEVLIHGNG"

RBS
3043. .3753
/standard_name="BIOD RBS"
3043. .3753
/gene="BIOD"
3043. .3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEVNTPTEVNGKTVASCALLOAKAAGYPRAGYKPVASGS
EKPEPELRNSDALAIORNSIOLDYATVPYTFAPPTSPHITSAGEPISLVASAG
LRALBOADWYLVGAGCGFTPLSDFTFADWVTOEQLVLLVGVKLGICINHAULTA
OVIQHGGLTLAGVANDVTPPKRHAETWTLITRMIPAPLLGELIPWLAENPENMATGK
YNLAFAVDASTLGFSTRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;

```

stem_loop      5583..5605
terminator     5583..5644
               /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
               TERMINATOR"
BASE COUNT     1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattattt 25
    |||||
Db 140 attgtcgcaagtcacagaattattt 164

RESULT 7
LOCUS      A38251      5872 bp      DNA
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL    LONZA AG (CH)
           Other publication PL 308301 950724
           Other publication CA 2145400 940414
           Other publication AU 4820293 940426
           Other publication HU 71781 960228
           Other publication SK 42095 951108
           Other publication CZ 9500809 950913
           Other publication FI 951547 950331
           Other publication JP 85016947 960227.
FEATURES
   source    1..5872
             /organism="Escherichia coli"
             /strain="DSM498"
             /db_xref="taxon:562"
             /clone="PB030A15-9"
             1141..1156
             /standard_name="BIOF RBS"
             1154..2308
             /gene="BIOF"
             1154..2308
             /gene="BIOF"
             /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
             /EC_number="2.3.1.47"
             /codon_start=1
             /transl_table=11
             /number=2
             /evidence=experimental
             /product="KAPA SYNTHASE"
             /protein_id="CAA02329.1"
             /db_xref="GI:2294850"
             /translation="MSWDEKINAAALDARRAADLRRTYPAVAGGAGRWLVADDPQYLN
             SSNDYLGISHHPQIRAMQCAEQFGIGSGSGVSVVHQALEELAEMLGTSFA
             LLFTSGFAANQAVTAAMAKEDRIADRLSHASLLEASLSPSOLRFAINDVTHLAR
             LLASPCQCOQAVTEGVGSMGDSAPLAELIQVYQOHNGWLMWDAGHGTGVIQEOGK
             SCWLQKVPKELIVTFPGKPGVSCAVICSTVADYLLQFARHLIYSTSMPAQQAQAI
             RASTAVIRSDGDAARRKIALITFRFAGVODLPFTLADSCSAIPLIVGDNSALQI
             AEKTRQSCVWTAIRPPTVPAGTARKLTLTAHEMODIRLLEVLHGNC"
RBS
3030..3045
/standard_name="BIOF RBS"
3043..3753
/gene="BIOF"

```

```

CDS
3043..3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/db_xref="GI:2294851"
/translation="MSKRYFVGTGDPVGVKTVASCALLQAKAGYRTAGYKPVASG
EKTEPGLRNSDALALQRNSSIQDDYATVNVPTAEPTSPHITSAGRPTEISIYMSAG
LRALEQQAQADWVLEVGAGGWFPLUSDFTFADWTOBOLPYILVYGVKLGGINAMLTA
QVIOHAGLTLAGWVANDVTPGKRHAERYMTTLTRMIPAPILGEPWLAENPENAAATGK
YINLAFVDASTLGTFTSRL"
BASE COUNT     1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattattt 25
    |||||
Db 140 attgtcgcaagtcacagaattattt 164

RESULT 8
LOCUS      A93674      5872 bp      DNA
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL    LONZA AG (CH)
           Other publication EP 0798384-A 1 01-OCT-1997;
           LONZA AG (CH)
           Location/Qualifiers
           1..5872
           /organism="Escherichia coli"
           /strain="DSM498"
           /db_xref="taxon:562"
           /clone="PB030A-15/9"
           1..96
           /function="PROMOTER PRAC"
           /evidence=experimental
           23..48
           /standard_name="PROMOTER PRAC"
           /evidence=experimental
           105..119
           /standard_name="BIOF RBS NO.9"
           /evidence=experimental
           117..1157
           /gene="BIOF"
           117..1157
           /gene="BIOF"
           /codon_start=1
           /transl_table=11
           /evidence=experimental
           /product="BIOTIN SYNTHASE"
           /protein_id="CAB69590.1"
           /db_xref="GI:6741863"
           /translation="MAHRRRWTLISOYTELEFKPRLDILFEAQOVYHROHFPDPPOVST
           LSIKTAGCPEDCKYCPQSSRYKTGLEARLMEVQVLESARKAKAAGSTRCGAAM

```

gene	HANRFLPDEIEOSLNGVYHOHHIOPITLMPDIALSAMRSKLGICATHLHBSROPRIL TRSOLORLQIAMPOOGGRPLTYHLFLGVIARE"
CDS	/number=1 /codon_start=1 /product="Biotin synthetase" /protein_id="AA023618.1" /db_xref="GI:145428" /translation="MSKRYFVTGTDTGKTAVSCALLQAARAGRYTAKYPVASGS EKRPEDLRNSDALALORNSSLQIDATVPYFAEPTSPHIIISAQEGPISLVNHSAG LRALDEOADVLEVGAGWETPLSDFTFADWTOELPYLLVAGKIGCIINHALITA QVQHAGGLTLAGVANDVTPPCKRHAHEMYTTLTRMIPRCMERSEFGLOKIKMRPES T"
BASE COUNT	1363 a 1554 c 1631 g 1245 t
ORIGIN	4626 bp upstream of HpaI site; 18 min on K-12 map.
Query Match	100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity	100.0%; Pred No. 0.091;
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	1 attctgcgaagtcacagaattatctt 25 Db 2035 ATTCTGCCAAGTCACAGAATTATT 2059
RESULT 6	A38246 5872 bp DNA PAT 05-MAR-1997
LOCUS	A38246 Sequence 1 from Patent WO9408023.
DEFINITION	A38246 A38246 GI:2294844
ACCESSION	A38246
VERSION	A38246.1
KEYWORDS	Escherichia coli. Bacterium. Organism
SOURCE	Escherichia coli. Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Bacteriella.
REFERENCE	1 (bases 1 to 5872) Birch,O., Brass,J., Fuhrmann,M. and Shaw,N. BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN Patent: WO 9408023-A 1 14-APR-1994;
AUTHORS	LONZA AG (CH)
TITLE	Other publication PL 308301 950724
JOURNAL	Other publication CA 2145400 940414 Other publication AU 4820293 940426 Other publication HU 71781 960228 Other publication SK 42095 951108 Other publication CZ 9500809 950913 Other publication FI 951547 950331 Other publication JP 85016947 960227. Location/Qualifiers 1..5872 /organism="Escherichia coli" /strain="DSM498" /db_xref="taxon:562" /clone="PB030A-15/9" 1..96 /function="PROMOTER PTAC" /evidence=experimental 23..28 /evidence=experimental 45..49 /standard_name="PROMOTER PTAC" /evidence=experimental 105..119 /standard_name="BIOB RBS NO.9" /evidence=experimental 117..1157 /gene="BIOB" 117..1157 /gene="BIOB"
COMMENT	
PROMOTER	
-35_signal	
-10_signal	
RBS	
gene	
CDS	

ACCESSION E00893
 VERSION GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hiroo, Y., Kojima, T. and Kimura, H.
 TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 COMMENT NIPPON SODA CO. LTD
 OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIRONO YOSHIMITSU, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
 C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT CDS 42..1079
 /product="biotin synthetase".

FEATURES
 source 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1121;
 Best Local Similarity 100.0%; Pred. NO. 0.088;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaatgcacagattattt 25
 |||||
 Db 65 attgtcgcaatgcacagattattt 89

RESULT 5
 ECOBIO 5793 bp DNA BCT 28-FEB-1994
 LOCUS E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
 DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
 protein, and dehydrobiotin synthetase (bioD), complete cds.
 J04423
 J04423.1 GI:145422
 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
 bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
 synthetase.
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and
 Johnson, O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 predicted
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
 A.Otsuka, 09-NOV-1988.
 FEATURES Location/Qualifiers

source 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=11
 /protein_id="AA023513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLRDGDKLPHRRHVENGMGVDGDNISPLANDDVPAGTK
 SEVATCYDPAAPGSGMHWVYVNLPAEDRYLPQGFSGGVAMPDGYLQTRTEGKGTG
 YDGAAPRGKGRHYFTYHALDIRIDVDEGASAAVNGFVNHHSLASASITAMFS"
 complement(633..1925)
 /gene="bioA"
 complement(633..1925)
 /codon_start=1
 /transl_table=11
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA023514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAFDQRHWHPTSMTPSLPYVYVVSABGCELLISDGR
 LVDGSSWMAAIGHYNHPOLNAAMKSQIDAMSHVMPFGITHAPAIELCRKIVASGRN
 ALECYFLADSGSVAVEVAMKALQYQWAKGARGRFLTERNGYIGDFFGASVCDPDN
 SMHSIMKGYLPENLFPAPAPQSRMDGEMDERMVGARLMAHREHIAVITPIVQGA
 GMRKTHPEMLKRIKICDREGILLIADEITAGRGRTGKIFACEHATLAPILLIGKA
 LTGGTWTLSATLITTRVETTSNCEAGCFHNGPTPKGNRPLACAAANSALIESGDMQ
 QOVADIEVQRLQAPARDAEMVADVRLVAGIVETTHPVNMAALQKFVEGQWIR
 PFGKLIYMPPIILIPOLRLTAVNRVAVODETFEQ"
 2012..3052
 /gene="bioB"
 2012..3052
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="AA023515.1"
 /db_xref="GI:145425"
 /translation="MAHRPRMTLSQVTELFKPLILDLFEAOQVYHROHFDPROVST
 LLSITGACPEQCKYCPTSRKTSLEERLMEVQYIESRAKKAAGSPCEGAM
 KNPHRDMPELEQWOGVAKMGLKCHNLTGLTSESQORLNAGIDTYNNHLDTSPE
 YGNIITTRVQERLDTLEKVRDAGIKVSGSIVIGETVKDRAGILQLANLPPES
 YPNIIMLVKGTPLADNDVDVDFIRITIAVARIMPTSVYVLSAGREOMEDYQWAK
 PMAGANSIFYGCKLLTTPNEEDKDLQFRKILGNPQOTAVLAGDNEQQLTEQALMT
 PPTDEVYAAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=11
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA023516.1"
 /db_xref="GI:145426"
 /translation="MSWQEKINAIDARRADALRRRPVAGAGRCMLVADROLYNF
 SSNDYIGSHHPQIIRAMQOAGDEFGISGSGSVVHQALEELAEALMGLSRA
 LFIISGFANQAVIIAAMAKEDRIADRLSHLSLEASLSPQLRRRAHNDVYHAR
 LIASPGQGOVVTEGVESMDGSAVLAETIOOVTOOHNGMILVDDAHGTGYIEGGRG
 SCWLOKVRPELLVYVPRGFGVGAVALCSSTVADYLLQFARHLITSMPAQAOL
 RASLAVIRSDGCDARREKIALITFRAGVQDLPPTLADSCSAIOTPLVGNRSRALQL
 AKRLQOQGWVAIRPVPVPACTARLRLITRAHEMODIRLLEVLHNG"
 4190..4945
 /gene="bioC"
 4190..4945
 /gene="bioC"
 /codon_start=1
 /transl_table=11
 /protein_id="AA023517.1"
 /db_xref="GI:145427"
 /translation="MATVNRKQAIIAAFGRAAAHYEQHADLQROSADALAMLPORRYT
 /HLDAGCGGMSRHRRHROKQVDTLDSPLVLPQGVAAKDDADHYTLADITSLPLATA
 TFDLAMSNI LAVOMCGLSTALRELYRVYRPGKVAFTLLVQGSLLPERRQANQVADERP

71 18.6 74.4 191562 2 AC087277 Homo sapi
72 18.6 74.4 192804 2 AC074289 Homo sapi
73 18.6 74.4 193295 2 AC073093 Homo sapi
74 18.6 74.4 193552 2 AC068463 Homo sapi
75 18.6 74.4 193561 2 AC090618 Homo sapi
76 18.6 74.4 194576 2 AC019147 Homo sapi
77 18.6 74.4 195165 8 ATCHRIV82 Homo sapi
78 18.6 74.4 196206 8 AC012368 Homo sapi
79 18.6 74.4 198402 8 ATCHRIV1 Homo sapi
80 18.6 74.4 198568 2 AC016754 Homo sapi
81 18.6 74.4 202851 2 AC015684 Homo sapi
82 18.6 74.4 206454 2 AC011227 Homo sapi
83 18.6 74.4 206504 2 AC012491 Homo sapi
84 18.6 74.4 228563 2 AC090219 Homo sapi
85 18.6 74.4 235985 2 AC068013 Homo sapi
86 18.6 74.4 253076 2 AC091905 Homo sapi
87 18.2 72.8 605 9 HSAPAF20 Homo sapi
88 18.2 72.8 25688 4 BTU15731 Homo sapi
89 18.2 72.8 38825 9 AC090098 Homo sapi
90 18.2 72.8 88958 9 AC008967 Homo sapi
91 18.2 72.8 104768 2 AC003023 Homo sapi
92 18.2 72.8 111312 9 AC011248 Homo sapi
93 18.2 72.8 118097 9 AL589826 Homo sapi
94 18.2 72.8 118673 9 HS1184F4 Human DNA
95 18.2 72.8 149646 2 AP003761 Oryza sat
96 18.2 72.8 152719 2 AC012545 Homo sapi
97 18.2 72.8 153087 2 AC017072 Homo sapi
98 18.2 72.8 165046 2 AC091991 Homo sapi
99 18.2 72.8 166471 2 AP003306 Homo sapi
100 18.2 72.8 172281 2 AC074200 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Paton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Source location/Qualifiers
1..1041
/organism="unknown"

BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattattc 25
|||||
Db 24 ATTGTCGCAAGTCACGAATTATTT 48

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Paton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Source location/Qualifiers
1..1041
/organism="unknown"

BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 attgtcgcaagtcacagaattattc 25
|||||
Db 24 ATTGTCGCAAGTCACGAATTATTT 48

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.

REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Escherichia.
FEATURES
Source Patent: GB 2216530-A 16 11-OCT-1989;
Location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="B10B"
24..1064
/gene="B10B"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490218"
/db_xref="SWISS-PROT:P12996"

/translation="MAHRRPMTLSOYTELFEKRLDLLEAQQVHRQHPDPOVST
LISIKTGACPEDCKTCPOSSRKTGLEERLMEVEVLESARRAKAGSTRCGAAM
KNFERDMPLLEQWVGKAMGLEACMTIGTISEQAOQLANAGLDYVNHNDTSPEF
YGNITTRYOERLDLLEKVRDAGIKVCSGIVIGETVDRAGLLDLQANLPTPES
VPIINMLVKVGTPLADNDVDVDAFDFRTIAVARIMPTSYVRLSGREOMNBOGAMC
FMAGANSIFYGCKLITTPNPDEDKDLQLRKRLGLNPGQAVAVLAGNEQQRLREQALMT
PDTDEVYNAAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattattc 25
|||||
Db 47 ATTGTCGCAAGTCACGAATTATTT 71

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:490218
KEYWORDS
SOURCE Unknown.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:54:31 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 attgcgcagctcacagaattattt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : GenEmbl : *

1: gb_da : *

2: gb_htg : *

3: gb_in : *

4: gb_om : *

5: gb_ov : *

6: gb_pat : *

7: gb_ph : *

8: gb_pl : *

9: gb_pr : *

10: gb_ro : *

11: gb_sts : *

12: gb_sy : *

13: gb_un : *

14: gb_vi : *

15: em_ba : *

16: em_fun : *

17: em_hum : *

18: em_in : *

19: em_om : *

20: em_or : *

21: em_ov : *

22: em_pat : *

23: em_ph : *

24: em_pl : *

25: em_ro : *

26: em_sts : *

27: em_sy : *

28: em_un : *

29: em_vi : *

30: em_htg_hum : *

31: em_htg_in : *

32: em_htg_rod : *

33: em_htg_hum : *

34: em_htg_in : *

35: em_htg_rod : *

36: em_htg_other : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	25	100.0	6 AR029499	AR029499 Sequence
2	25	100.0	6 AR034916	AR034916 Sequence
3	25	100.0	6 A11530	A11530 B10b gene o
4	25	100.0	6 E00893	E00893 Genomic DNA
5	25	100.0	6 E00893	J04423 E.coli 7,8-
6	25	100.0	6 A38246	A38246 Sequence 1
7	25	100.0	6 A38251	A38251 Sequence 6
8	25	100.0	6 A93674	A93674 Sequence 1
9	25	100.0	6 A93679	A93679 Sequence 6
10	25	100.0	6 AR101809	AR101809 Sequence
11	25	100.0	6 AR101810	AR101810 Sequence
12	25	100.0	6 AE000180	AE000180 Escherich
13	25	100.0	6 AE005258	AE005258 Escherich
14	25	100.0	6 AP002553	AP002553 Escherich
15	24	96.0	1 AF250776	AF250776 Unculture
16	21.8	87.2	1 AF248314	AF248314 Unculture
17	20.4	81.6	1 AF250770	AF250770 Unculture
18	20.2	80.8	2 AC025510	AC025510 Homo sapi
19	20.2	80.8	2 AC020741	AC020741 Homo sapi
20	19.4	77.6	2 AC084797	AC084797 Homo sapi
21	19.2	76.8	2 HS82311	HS82311 Human DNA s
22	19.2	76.8	2 AC020546	AC020546 Homo sapi
23	19.2	76.8	2 CNS057C0	CNS057C0 Human chr
24	19.2	76.8	2 AC022047	AC022047 Homo sapi
25	19.2	76.8	2 CNS07ED4	CNS07ED4 Human chr
26	19.2	76.8	2 AC025062	AC025062 Homo sapi
27	19.2	76.8	2 AC007564	AC007564 Homo sapi
28	19.2	76.8	2 AC026977	AC026977 Homo sapi
29	19.2	76.8	2 AC020772	AC020772 Homo sapi
30	19.2	76.8	2 AC004470	AC004470 Homo sapi
31	19.2	76.8	2 AC007553	AC007553 Homo sapi
32	18.8	75.2	2 A1133462	A1133462 Human DNA
33	18.8	75.2	2 AC012380	AC012380 Genomic S
34	18.8	75.2	2 AC064870	AC064870 Homo sapi
35	18.6	74.4	2 A75959	A75959 Sequence 1
36	18.6	74.4	2 ATRTRIG	X66016 A. thaliana
37	18.6	74.4	2 AE007865	AE007865 Clostridi
38	18.6	74.4	2 CEF19C6	Z48006 Caenorhabdi
39	18.6	74.4	2 AFI28457	AFI28457 Oryza sat
40	18.6	74.4	2 HS45P1	AL035397 Human DNA
41	18.6	74.4	2 ATM4E13	AL022023 Arabidops
42	18.6	74.4	2 AC024419	AC024419 Homo sapi
43	18.6	74.4	2 ATR1235	AL035522 Arabidops
44	18.6	74.4	2 AC016406	AC016406 Homo sapi
45	18.6	74.4	2 AC016406	AC016406 Homo sapi
46	18.6	74.4	2 ATRP22K18	AL035386 Arabidops
47	18.6	74.4	2 AC019226	AC019226 Homo sapi
48	18.6	74.4	2 AC034229	AC034229 Homo sapi
49	18.6	74.4	2 A1357552	AL357552 Homo sapi
50	18.6	74.4	2 AC016743	AC016743 Homo sapi
51	18.6	74.4	2 AC023220	AC023220 Homo sapi
52	18.6	74.4	2 AC027141	AC027141 Homo sapi
53	18.6	74.4	2 AC015953	AC015953 Homo sapi
54	18.6	74.4	2 AC069411	AC069411 Homo sapi
55	18.6	74.4	2 AC011155	AC011155 Homo sapi
56	18.6	74.4	2 AC022371	AC022371 Homo sapi
57	18.6	74.4	2 AC026000	AC026000 Homo sapi
58	18.6	74.4	2 AC019278	AC019278 Homo sapi
59	18.6	74.4	2 AC092668	AC092668 Homo sapi
60	18.6	74.4	2 AC025240	AC025240 Homo sapi
61	18.6	74.4	2 AC022309	AC022309 Homo sapi
62	18.6	74.4	2 AC087695	AC087695 Homo sapi
63	18.6	74.4	2 AC091647	AC091647 Homo sapi
64	18.6	74.4	2 A1596327	AL596327 Homo sapi
65	18.6	74.4	2 AP001934	AP001934 Homo sapi
66	18.6	74.4	2 AC018515	AC018515 Homo sapi
67	18.6	74.4	2 AC068595	AC068595 Homo sapi
68	18.6	74.4	2 AC023180	AC023180 Homo sapi
69	18.6	74.4	2 AC041009	AC041009 Homo sapi
70	18.6	74.4	2 AC036131	AC036131 Homo sapi

Db 112 GACATTTTCACAGCCACAGAAATT 135

RESULT 45

BI093008/c

LOCUS 602858287F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:499508 5',
EST 20-JUN-2001

DEFINITION mRNA sequence.

ACCESSION BI093008 GI:14511338

VERSION BI093008.1

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 838)
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM11029 Row: n Column: 21
High quality sequence stop: 740.

FEATURES

source

1..838
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:499508"
/clone_lib="NIH-MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

BASE COUNT 225 a 141 c 167 g 305 t
ORIGIN

Query Match

Best Local Similarity 83.3%; Score 17.6; DB 11; Length 838;
Matches 20; Conservative 0; Mismatches 4; Indels 0;

OY 2 acatgtgcgaagtcacagaatta 25
||||| |||| || ||||
Db 697 ACATGTGTGCGACAGCCAGAAATGA 674

Search completed: December 26, 2001, 12:11:03
Job time: 6425 sec

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 661;
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acattgtcgaagtcacagaatta 25
 ||||||| || |||||||
 Db 221 ACATTGCTCTAAATTTCACAGATT 198

RESULT 43

AZ289880 667 bp DNA GSS 27-JUL-2000
 LOCUS RPCI-23-129G2.TJ RPCI-23 Mus musculus genomic clone RPCI-23-129G2,
 DEFINITION DNA sequence.

ACCESSION AZ289880
 VERSION AZ289880.1 GI:9531666

KEYWORDS

GSS.
 house mouse.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 667)

AUTHORS

Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinet,
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
 and Fraser, C. M.

TITLE

Mouse BAC End Sequences from Library RPCI-23

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (http://www.tigr.org/tldb/bac-ends/mouse/bac_end_intro.html)
 Plate: 129 row: G column: 2
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..667
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-129G2"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1;
 EcoRI, Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI methylase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 192 a 149 c 169 g 157 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 667;
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acattgtcgaagtcacagaatta 25
 ||||||| || |||||||
 Db 180 ACATTGCTCTAAATTTCACAGATT 157

RESULT 44

AZ334568 719 bp DNA GSS 29-SEP-2000
 LOCUS 1M0064F02F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
 DEFINITION clone UUGCLM0064F02 F, DNA sequence.

ACCESSION AZ334568
 VERSION AZ334568.1 GI:10402021

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 719)

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relliy,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0064 row: F column: 02
 Seq primer: CGTTGTAACGACGCGCACT
 Class: plasmid ends
 High quality sequence stop: 719.

FEATURES

source

Location/Qualifiers

1..719
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGCLM0064F02"
 /clone_lib="Mouse 10kb plasmid UUGCLM library"
 /sex="Male"
 /lab_host="E. coli strain XL10-gold, m1-resistant, F-"
 /note="Vector: pMD42uv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 230 a 111 c 108 g 270 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 719;
 Best Local Similarity 83.3%; Pred. No. 5.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaattgtcgaagtcacagaatt 24
 ||||||| || || |||||||

QY	DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	JOURNAL	COMMENT	REFERENCE	AUTHORS
1	421	GC0ATTTGAAGCAAGACACGAAATT 444	661 bp DNA	AO389869	RPC11-142A6	TV	RPC1-11	Homo sapiens genomic clone	RPC1-11-142A6, DNA sequence.	AO389869	AO389869	GI:4360892	GSS.
24								human.	Use of BAC end Sequences from Library RPC1-11 for Sequence-Ready Map Building	Unpublished (1997)	Other-GSSs: RPC11-142A6.TJ	Contact: Shaying Zhao, William Nierman, Mark Adams	Department of Eukaryotic Genomics
													The Institute for Genomic Research
													9712 Medical Center Dr., Rockville, MD 20850
													Tel: 301 838 0200
													Fax: 301 838 0208
													Email: hbett@tigr.org
													Clones are derived from the human BAC library RPC1-11. For BAC
													library availability, please contact Pieter de Jong
													(pieter@tigr.org, med.buffalo.edu). Clones may be purchased from
													BACAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
													Research Genetics (info@resgen.com). BAC end search page:
													http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
													Seq primer: 175
													Class: BAC
													Location/Qualifiers
													1. 661
													/organism="Homo sapiens"
													/db_xref="GDB:7554149"
													/db_xref="taxon:9606"
													/clone="RPC1-11-142A6"
													/clone_lib="RPC1-11"
													/sex="Male"
													/cell_type="Lymphocytes"
													/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Site_3: EcoRI; Human Male BAC Library"
													244 a 109 c 114 g 194 t

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114[gb]AF12902.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 138 a 127 c 122 g 140 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 527;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacatgcgaagtcacgaatt 24
||||| ||||| ||||| ||||| ||
Db 385 GACATTTCCGCGACGACACGCTTT 408

RESULT 36
BI210793 538 bp mRNA EST 11-JUL-2001
LOCUS EST58833 cTOS Lycopersicon esculentum cDNA clone cTOS2418 5' end,
DEFINITION mRNA sequence.
ACCESSION BI210793.1 GI:1468517
VERSION BI210793.1
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 538)
van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,
Romling, C. and Tanksley, S.
Generation of ESTs from Tomato Suspension Cultures
Unpublished (2001)
Contact: CGEI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
Location/Qualifiers

FEATURES
source

1. 538
/organism="Lycopersicon esculentum"
/cultivar="TA496, E6203"
/db_xref="taxon:4081"
/clone="cTOS2418"
/clone_lib="cTOS"
/tissue_type="suspension cultures"
/lab_host="SOLR"
/note="Vector: Bluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Suspension cultures of L.esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 15%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."
BASE COUNT 164 a 96 c 116 g 162 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 538;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacatgcgaagtcacgaatt 24
||||| ||||| ||||| ||||| ||
Db 300 GCCATTGACGACGACACGAGATT 323

RESULT 37
A2476585 599 bp DNA GSS 04-OCT-2000
LOCUS A2476585
DEFINITION IM0295H03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0295H03 R. DNA sequence.
ACCESSION A2476585
VERSION A2476585.1 GI:10634710
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 599)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly
, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0295 row: H column: 03
Seq primer: CACACAGGAAACACACCTATGACC
Class: plasmid ends
High quality sequence stop: 599.
Location/Qualifiers

FEATURES
source

1. 599
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0295H03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
<http://www.jax.org/resources/documents/dnares/>. The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g14732114[gb]AF12902.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent *E. coli* XL10-Gold (Stratagene) cells
and selected for ampicillin resistance.
BASE COUNT 179 a 115 c 102 g 203 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 599;

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 499)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3157 row: H column: 15
Class: BAC ends
High quality sequence stop: 499.
Location/Qualifiers
1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-coli DH10B"

BASE COUNT 122 a 80 c 80 g 216 t 1 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 499;
Best Local Similarity 83.3%; Pred. No. 5,1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgaagtcacagaatta 25
||||| ||||||| |||
Db 206 ACATTGTACAGTCACATATATA 183

RESULT 34
LOCUS A0215052 517 bp DNA GSS 18-SEP-1998
DEFINITION HS.3244.A2.C03.T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3244 Col=6 Row=E, DNA sequence.
ACCESSION A0215052
VERSION A0215052.1 GI:3626253
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 517)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector

Plate: 3244 row: E column: 6
Class: BAC ends
High quality sequence stop: 517.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-coli DH10B"

BASE COUNT 118 a 87 c 92 g 220 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 517;
Best Local Similarity 83.3%; Pred. No. 5,1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgaagtcacagaatta 25
||||| ||||||| |||
Db 254 ACATTGTACAGTCACATATATA 231

RESULT 35
LOCUS A2337152 527 bp DNA GSS 29-SEP-2000
DEFINITION IM0067B09R Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG1M0067B09 R, DNA sequence.
ACCESSION A2337152
VERSION A2337152.1 GI:10407000
KEYWORDS GSS.
SOURCE Mus musculus.
ORGANISM Mus musculus
REFERENCE Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus. 1 (bases 1 to 527)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacom,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0067 row: B column: 09
Seq primer: CACACGACAAACAGCATGACC
Class: Plasmid ends
High quality sequence stop: 527.
Location/Qualifiers
1..527
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUCG1M0067B09"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html

Seq primer: 17
Class: BAC ends.

FEATURES

Location/Qualifiers

1. 403
/organism="Homo sapiens"
/db_xref="GDB:7534293"
/db_xref="taxon:9606"
/clone="RPC1-11-142G6"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: PBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"

BASE COUNT

137 a 59 c 69 g 138 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 403;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acatgtcgcaagtcacagaatta 25

Db 194 ACATTGCTAATTTCAGAGATTA 171

RESULT 31

LOCUS

AO314476 431 bp DNA GSS 04-MAY-1999
RPC11-104B8.TV RPC1-11 Homo sapiens genomic clone RPC1-11-104B8,
DNA sequence.

ACCESSION AO314476 GI:4045939

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 431)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Sun,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: RPC11-104B8.TV
Contact: Sheng Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pletier de Jong
(pletier@jeng.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends

FEATURES

SOURCE

ORIGIN

Location/Qualifiers
1. 431
/organism="Homo sapiens"
/db_xref="GDB:7539583"
/db_xref="taxon:9606"
/clone="RPC1-11-104B8"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: PBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"

BASE COUNT

137 a 87 c 101 g 106 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 431;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acatgtcgcaagtcacagaatta 25

Db 161 ACATTGCTAATTTCAGAGATTA 184

RESULT 32

LOCUS

AL040883/c 460 bp mRNA EST 29-FEB-2000
DKFZp434H0915.t1.434 (synonym: htes3) Homo sapiens cDNA clone
DKFZp434H0915.5', mRNA sequence.

ACCESSION AL040883 GI:5409828

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 460)
Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
EST (Bloecker, et al.)
Unpublished (1999)
Contact: Bloecker H
MIPS
Am Klopferplatz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No SI sequence available.
This clone (DKFZp434H0915) is available at the RZPD in Berlin
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

SOURCE

Location/Qualifiers
1. 460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434H0915"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPori1; Site_1: NotI; Site_2: SalI"

BASE COUNT

159 a 69 c 125 g 107 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 460;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacatgtcgcaagtcacagaatt 24

Db 226 GACATTGTAGCACATCATGAAATT 203

RESULT 33

LOCUS

AO142135 499 bp DNA GSS 24-SEP-1998
HS 3157.B1.D08.T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone plate=3157 Col-15 Row=H, DNA sequence.

ACCESSION AO142135 GI:3532788

VERSION

KEYWORDS


```

/clone="UI-R-B01-aps-a-08-0-UI"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratelst.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-B01
TAG_TISSUE=mid-brain
TAG_SEQ=CATGG

```

```

Query Match      71.2%  Score 17.8; DB 10; Length 382;
Best Local Similarity 90.5%  Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      3  catgtgcgaagtcacagaat 23
        ||| ||| ||| ||| ||| |||
Db      50  CATGTGCGCAAGTCTCACAGT 70

```

```

RESULT 24
AWB33543 194 bp mRNA EST 18-MAY-2000
LOCUS OVA-TT00008-091199-026-c05 TT00008 Homo sapiens cDNA, mRNA sequence.
DEFINITION AWB33543
ACCESSION AWB33543
VERSION AWB33543.1 GI:7927517
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 194)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-TT00008-091
199-026-c05&tl=1999-11-09&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 194.
Location/Qualifiers
1. 194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT00008"

```

```

FEATURES
source
1. 194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT00008"

```

```

FEATURES
source
1. 194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT00008"

```

```

/dev_stage="Adult"
/note="Organ: testis; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```

```

Query Match      70.4%  Score 17.6; DB 10; Length 194;
Best Local Similarity 83.3%  Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      1  gacatgtgcgaagtcacagaat 24
        ||| ||| ||| ||| ||| |||
Db      154  GTCACTGTGCGCAAGTCCCAAGATT 131

```

```

RESULT 25
AWB33554 203 bp mRNA EST 18-MAY-2000
LOCUS OVA-TT00008-091199-026-e10 TT00008 Homo sapiens cDNA, mRNA sequence.
DEFINITION AWB33554
ACCESSION AWB33554
VERSION AWB33554.1 GI:7927528
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 203)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-TT00008-091
199-026-e10&tl=1999-11-09&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 203.
Location/Qualifiers
1. 203
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT00008"
/dev_stage="Adult"
/note="Organ: testis; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```

```

BASE COUNT      36 a      56 c      43 g      67 t      1 others
ORIGIN

```

JOURNAL
COMMENT

Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamads@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@ed.jong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: 77
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..352
/organism="Homo sapiens"
/db_xref="GDB:7526293"
/db_xref="taxon:9606"
/clone="RPCI-11-69H14"
/clone_11b="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI; RPC11 Human Male BAC Library"
BASE COUNT 86 a 91 c 98 g 75 t 2 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 352;
Best Local Similarity 90.5%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 acatgtcgaagtcacaga 22
11111111111111111111
Db 293 ACATGCTCAGTCACAGCA 313

RESULT 22
R21156/c 379 bp mRNA EST 18-APR-1995
LOCUS yh20a04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
DEFINITION IMAGE:130254.5', mRNA sequence.
R21156
ACCESSION R21156.1 GI:775937
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 379)

REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marrs, M., Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
CONTACT: Wilson RK
JOURNAL Washington University School of Medicine
MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 987
High quality sequence stops: 337
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 987 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 337.

FEATURES
source

Location/Qualifiers
1..379
/organism="Homo sapiens"
/db_xref="GDB:537603"
/db_xref="taxon:9606"
/clone="IMAGE:130254"
/clone_11b="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="PH10 (ampicillin resistant)"
/note="Organ: placenta; Vector: p773D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' ACTGGAGAGATTCGCGCGGCGAGATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo."
BASE COUNT 123 a 56 c 83 g 117 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 379;
Best Local Similarity 90.5%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ttgtgcgaagtcacagaatta 25
11111111111111111111
Db 325 TTGTGCCAGTCACACAAATTA 305

RESULT 23
BE101686 382 bp mRNA EST 13-JUN-2000
LOCUS UI-R-B01-aps-a-08-0-UI.s1 UI-R-B01 Rattus norvegicus cDNA clone
DEFINITION UI-R-B01-aps-a-08-0-UI 3', mRNA sequence.
BE101686
ACCESSION BE101686.1 GI:8493784
VERSION EST.
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 382)

REFERENCE Bonaldo, M.F., Lennon, G. and Soares, M.B.
AUTHORS Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iue.wiueg.iowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized mid-brain library cDNA library preparation; M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..382
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

Human PAC End sequences for sequence-ready map building

TITLE	JOURNAL	COMMENT
R., Waterston, R. and Wilson, R. Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com		High quality sequence stop: 375.

FEATURES	Location/Qualifiers
source	1. .388

BASE COUNT	123 a	67 c	92 g	106 f
ORIGIN				

Query Match	72.8 %;	score 18.2;	DB 10;	length 388;
Best Local Similarity	87.0 %;	Pred. No. 2.6e+02;		
Matches 20; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```
QY      3 catgtgcgaagtcacagaatcta 25
          ||| ||||| ||||| |||||
Db      328 CATACACCGCAAGTCACAGATTTA 350
```

RESULT	13			
T83852				
LOCUS	T83852	444 bp	mRNA	EST ¹
DEFINITION	yd6db07.s1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone			16-MAR-1995

ACCESSION	T83852
VERSION	T83852.1
KEYWORDS	GI:712140
SOURCE	EST.
ORGANISM	human.
	tissue: cationic

ORGANISM
Homo sapiens
Eukaryota: Metazoa; Chordata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
REFERENCE
1 (bases 1 to 444)
AUTHORS
Hiller, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisson, E., Waterston

TITLE	, R., Williamson A., Wohldmann, P. and Wilson, R.
JOURNAL	The Nashu-Merck EST Project
COMMENT	Unpublished (1995)
	Contact: Wilson RK
	Washington University School of Medicine

Washington University School of Medicine
4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 939
High quality sequence stops: 334 Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 939 Std Error: 0.00
Seq primer: T3
High quality sequence stop: 334.

```

FEATURES
source
location/Qualifiers
1. .444
General name="Hume 3311000"

```

BASE COUNT ORIGIN	108 a	69 c	100 g	162 t	5 others
----------------------	-------	------	-------	-------	----------

Query Match	72.8%;	Score 18.2;	DB 11;	Length 444;
Best Local Similarity	87.0%;	Pred. No. 2.6e+02;		
Matches 20; Conservative		Mismatches 3;	Indels 0;	Gaps 0

```
QY      2  acattgtcgcaagtcacagaatt 24
          | ||||| ||||| |||||
Db     188 ATATTGTTGCAAGTACAGATT 210
```

LOCUS	DEFINITION	RESULT 14
N58785		
N58785	454 bp mRNA	EST
		30-JAN-1997
	yy76c09.s1 Scores fetal liver spleen INF15 Homo sapiens cDNA clone	
	IMAGE:248656.3; similar to contans Alu repetitive element:contans	
	element LI repetitive element ;, mRNA sequence.	

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE	1 (bases 1 to 454)
AUTHORS	Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., Dubucq, T., Fagello, A., Gish, W., Hawkins, M., Hiltman, K., Kucaba, T., Laquey, M., Le, N., Madis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellendy, G., Soares, M.B., Watson, R., Wilson, R. and Mairra, M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	Contact: Wilson RK Washington University School of Medicine

and blunting of cDNA termini. EcoRI adapters were ligated, followed by kinase treatment and digestion with XhoI.

The cDNAs were then size-fractionated using Sephacryl S-500 column and then ligated into EcoRI and XhoI digested Lambda Zap Express vector. The ligation product was packaged using Gigapack II packaging extract. The library had primary titre of approx. 1x10⁶. Clones from the primary library were randomly selected for single pass sequencing."

BASE COUNT 108 a 53 c 64 g 77 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 302;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaatt 24
||||| ||||| |||||
DB 209 ACATTGTCCAGTCACAGATT 187

RESULT 8 B1049694 317 bp mRNA EST 15-JUN-2001
LOCUS B1049694 CM2-GN0288-020101-667-d05 GN0288 Homo sapiens cDNA, mRNA sequence.
ACCESSION B1049694
VERSION B1049694.1 GI:14457224
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS 1 (bases 1 to 317)
Dias Neto, F., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&cl2=CM2-GN0288-020101-667-d05&cl3=2001-01-02&cl4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 317.
Location/Qualifiers
1..317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0288"
/dev_stage="Adult"
/note="Organ: Placenta, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 94 a 71 c 46 g 106 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 317;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaatt 24
||||| ||||| |||||
DB 184 ATATTGTCCAGTCACAGATT 206

RESULT 9 AL042323/c 357 bp mRNA EST 29-FEB-2000
LOCUS AL042323 DKEP43401120.F1.434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKEP43401120.5', mRNA sequence.
ACCESSION AL042323
VERSION AL042323.2 GI:5866743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS 1 (bases 1 to 357)
Oltenevaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
EST (Oltenevaelder, et al.)
Unpublished (1999)
On Jul 8, 1999 this sequence version replaced gi:5421665.
COMMENT Contact: Oltenevaelder B
MPS
Am Kiofierspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
This clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medicomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available.
This clone (DKFZ43401120) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZ43401120"
/clone_id="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

FEATURES

BASE COUNT 132 a 74 c 60 g 91 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 357;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaatt 24
||||| ||||| |||||
DB 206 ACATTGTCCAGTCACAGATT 184

RESULT 10 BE800706 357 bp mRNA EST 20-SEP-2000
LOCUS BE800706 sg97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-c1049-1091 5' similar to TR:082150 O82150 CHLOKROPLAST FTSH
PROTEASE.; mRNA sequence.
ACCESSION BE800706
VERSION BE800706.1 GI:10231818

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 564)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10Kb plasmid inserts

TITLE Unpublished (2000)
JOURNAL Contact: Robert B. Weiss
COMMENT University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0284 row: A column: 02
Seq primer: CACACAGGAACAGCATGACGC
Class: plasmid ends
High quality sequence stop: 564.

FEATURES
source
1..564
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U00C2M0284A02"
/clone_1lib="Mouse 10kb plasmid U00C2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b/AF12972.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 175 a 153 c 103 g 133 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 564;
Best Local Similarity 84.0%; Pired. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatta 25
||||||| | ||||| |||||
Db 429 GACATTGTCCAGGTCCACAGACTA 453

RESULT 4
AZ083812 596 bp DNA GSS 08-MAY-2000
LOCUS
DEFINITION RPCI-23-22N4.TJ RPCI-23 Mus musculus genomic clone RPCI-23-22N4,
ACCESSION A2083812
VERSION A2083812.1 GI:7725545
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 596)
Zhao, S., Nieman, W., Feldlyum, T., Malek, J., Shatsman, S., Aklnret, B., Levins, M., McGann, S., Isegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)

TITLE Unpublished (1999)
JOURNAL Contact: Shaying Zhao
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buhalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buhalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 22 row: N column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..596
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-22N4"
/clone_1lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1; EcoRI: Site 2; EcoRI: Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL life Technologies)."

BASE COUNT 174 a 108 c 95 g 219 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 596;
Best Local Similarity 84.0%; Pired. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatta 25
||||||| | ||||| |||||
Db 561 GGCATTGTCCAAAGTCACAGACTA 585

RESULT 5
AM155118 869 bp mRNA EST 03-NOV-1999
LOCUS
DEFINITION mg1e0002K13f Rice blast infection stage cDNA library Oryza sativa/Pyricularia oryzae mixed EST library cDNA clone mg1e0002K13f, mRNA sequence.
ACCESSION AM155118
VERSION AM155118.1 GI:6223987
KEYWORDS
SOURCE Oryza sativa/Pyricularia oryzae mixed EST library.
ORGANISM Oryza sativa/Pyricularia oryzae mixed EST library.
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 869)
AUTHORS Raunaree, R., Choi, W. and Dean, R.A.
TITLE Identification and characterization of genes expressed by the rice blast pathogen and rice during infection stage
JOURNAL Unpublished (1999)
COMMENT Contact: Dean, R.A.

```

c 86 17 68.0 432 11 284075 284075 SSZ84075 PO
c 87 17 68.0 435 11 BF750003 BF750003 CM3-BND38
c 88 17 68.0 448 13 A2260776 A2260776 RPT-23-1
c 89 17 68.0 448 13 A2928134 A2928134 479.d1f05
c 90 17 68.0 466 13 A0711120 A0711120 HS-5367_B
c 91 17 68.0 497 11 BF606242 273224 MA BF606242 273224 MA
c 92 17 68.0 518 13 A2928522 479.d1f10 A2928522 479.d1f10
c 93 17 68.0 523 13 A284402 RPT-23-1 A284402 RPT-23-1
c 94 17 68.0 526 10 A1443958 A1443958 sa28f03.y
c 95 17 68.0 530 13 TA374C050 AL494389 T. brucei
c 96 17 68.0 532 10 BE645662 BE645662 7e74c06.x
c 97 17 68.0 547 10 AW348119 AW348119 GM210001A
c 98 17 68.0 552 13 A2986136 A2986136 2M0268B1A
c 99 17 68.0 568 13 A0724791 A0724791 HS-5381_A
c 100 17 68.0 570 13 A0677763 A0677763 HS-5516_A

```

ALIGNMENTS

```

RESULT 1
LOCUS A071996/c
DEFINITION HS_5413_B2.F07.SP6E RPT-11 Human Male BAC Library Homo sapiens
ACCESSION A071996
VERSION A071996.1 GI:5651724
KEYWORDS GSS.
SOURCE genomic clone Plate=989 Col=14 Row=L, DNA sequence.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

```

```

REFERENCE 1
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPT-11. For BAC
library availability, please contact Pieter de Jong
(pjeter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 989 row: L column: 14
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 576.

```

FEATURES

```

source
1..576
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=989 Col=14 Row=L"
/clone_lib="RPT-11 Human Male BAC Library"
/sex="Male"
/note="Vector: pBAC3.6; Site:1: EcoRI; Site:2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT 167 a 126 c 109 g 168 t 6 others
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 13; Length 576;
Best Local Similarity 87.5%; Pred. No. 95;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 acatgtcgaagtcacagaatta 25
||||| ||| ||| |||
Db 219 ACATTGTCTCAAGCAGACACTTA 196

```

```

RESULT 2
LOCUS A0335616/c
DEFINITION HS_5015_B2.C08.SP6E RPT11 Human Male BAC Library Homo sapiens
ACCESSION A0335616
VERSION A0335616.1 GI:4142595
KEYWORDS GSS.
SOURCE genomic clone Plate=591 Col=16 Row=F, DNA sequence.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

```

```

REFERENCE 1
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 591 row: F column: 16
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 494.

```

FEATURES

```

source
1..494
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=591 Col=16 Row=F"
/clone_lib="RPT11 Human Male BAC Library"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; RPT11 Human Male BAC Library"
BASE COUNT 207 a 82 c 91 g 112 t 2 others
ORIGIN

```

```

Query Match 74.4%; Score 18.6; DB 13; Length 494;
Best Local Similarity 84.0%; Pred. No. 18e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 gacatgtcgaagtcacagaatta 25
||||| ||| ||| |||
Db 373 GACATTGTCTCAAGCAGACTTA 349

```

```

RESULT 3
LOCUS A2997800
DEFINITION 2M0284A02R Mouse 10kb plasmid U0GC2M library Mus musculus genomic
ACCESSION A2997800
VERSION A2997800.1 GI:13869027
KEYWORDS GSS.
SOURCE house mouse.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:23:58 : Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-1
Perfect score: 25
Sequence: 1 gacatgctgcagctcacagatla 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estfun:*
2: em_estlin:*
3: em_estlin:*
4: em_estlin:*
5: em_estlin:*
6: em_estlin:*
7: em_estlin:*
8: em_estlin:*
9: em_estlin:*
10: em_estlin:*
11: em_estlin:*
12: em_estlin:*
13: em_estlin:*
14: em_estlin:*
15: em_estlin:*
16: em_estlin:*
17: em_estlin:*
18: em_estlin:*
19: em_estlin:*
20: em_estlin:*
21: em_estlin:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	19.2	76.8	576	13	A0771996 HS_5413.B
C 2	18.6	74.4	494	13	A0355616 HS_5015.B
C 3	18.6	74.4	564	13	A2987800 2M0284A02
4	18.6	74.4	596	13	A2083812 RPTC1-23-2
5	18.6	74.4	869	10	AW155118 mgie0002K
6	18.2	72.8	169	11	W18068 mb85d03.r1
7	18.2	72.8	302	11	BF172143 PCIA225.M
8	18.2	72.8	317	11	BI049694 CM2-GN028
C 9	18.2	72.8	357	10	AL042323 DKFZP434O
C 10	18.2	72.8	357	11	BE800706 sg97c06.Y
C 11	18.2	72.8	387	13	BE800706 sg97c06.Y
C 12	18.2	72.8	388	10	AW781150 sl89906.Y

13	18.2	72.8	444	11	T83852
14	18.2	72.8	454	11	N58785
15	18.2	72.8	506	10	AI017333
16	18.2	72.8	557	13	A0437316
17	18.2	72.8	567	13	A225990 RPTC1-23-5
18	18.2	72.8	622	13	BI125940
C 19	18.2	72.8	709	13	A020375
C 20	18.2	72.8	751	10	BE68717
C 21	17.8	71.2	352	13	A0238325
C 22	17.8	71.2	379	11	BE101686
C 23	17.8	71.2	382	11	R21156
C 24	17.6	70.4	194	10	AM833543
C 25	17.6	70.4	203	10	AM833543
C 26	17.6	70.4	225	10	AM057146
C 27	17.6	70.4	232	10	AM833631
C 28	17.6	70.4	232	10	AM833631
C 29	17.6	70.4	274	10	AM833631
C 30	17.6	70.4	403	13	A0385232
C 31	17.6	70.4	431	13	A0314476
C 32	17.6	70.4	460	10	AL040883
C 33	17.6	70.4	499	13	A0142135
C 34	17.6	70.4	517	13	A0215052
C 35	17.6	70.4	527	13	A2337152
C 36	17.6	70.4	538	11	BI210793
C 37	17.6	70.4	559	13	A2476585
C 38	17.6	70.4	609	13	A2325063
C 39	17.6	70.4	615	13	A0602696
C 40	17.6	70.4	643	13	A0306368
C 41	17.6	70.4	659	13	AM038199
C 42	17.6	70.4	661	13	A0389869
C 43	17.6	70.4	667	13	A2289880
C 44	17.6	70.4	719	13	A2343568
C 45	17.6	70.4	838	11	BI093008
C 46	17.6	70.4	996	13	CNS07CXL
C 47	17.6	70.4	1011	13	CNS07DOY
C 48	17.4	69.6	531	13	A0806020
C 49	17.4	69.6	612	13	A0403329
C 50	17.4	69.6	622	13	A2385697
C 51	17.2	68.8	163	10	AV006577
C 52	17.2	68.8	317	13	A2456884
C 53	17.2	68.8	321	13	D12416
C 54	17.2	68.8	376	13	A0682180
C 55	17.2	68.8	379	10	AA920661
C 56	17.2	68.8	390	10	BE060601
C 57	17.2	68.8	423	10	AA398717
C 58	17.2	68.8	440	11	R933225
C 59	17.2	68.8	441	11	H02904
C 60	17.2	68.8	451	10	AA448276
C 61	17.2	68.8	483	10	AM202805
C 62	17.2	68.8	499	10	AA184553
C 63	17.2	68.8	531	13	A0401672
C 64	17.2	68.8	543	13	A0977126
C 65	17.2	68.8	551	13	A2072579
C 66	17.2	68.8	622	11	EG779264
C 67	17.2	68.8	677	13	A2650467
C 68	17.2	68.8	702	11	BI222905
C 69	17.2	68.8	751	10	BE394042
C 70	17.2	68.8	779	10	BE613352
C 71	17.2	68.8	849	10	BE548526
C 72	17.2	68.8	851	11	BE957936
C 73	17.2	68.8	1015	11	BE966737
C 74	17.2	68.8	1101	11	BG617252
C 75	17.2	68.8	197	10	AV357339
C 76	17.2	68.8	231	13	A2481861
C 77	17.2	68.8	250	13	A2492553
C 78	17.2	68.8	271	11	N81375
C 79	17.2	68.8	289	13	A0070825
C 80	17.2	68.8	366	11	Z14476
C 81	17.2	68.8	382	10	AI938247
C 82	17.2	68.8	396	10	AU022999
C 83	17.2	68.8	406	10	AA787570
C 84	17.2	68.8	419	10	BE068478
C 85	17.2	68.8	430	11	BE081507

T83852 yd66b07.s1
N58785 yv76c09.s1
AI017333 ov31h07.x
A0437316 HS_5083.A
A225990 RPTC1-23-5
BI125940 RPTC1-24-3
A020375 CIT-HSP-2
BE68717 CM700007A
A0238325 RPTC11-69
R21156 yh20a04.r1
BE101686 UI-R-B01-
AM833543 OV4-TT000
AM833543 OV4-TT000
AM057146 CA01605.Y
AM833631 OV4-TT000
AM833631 OV4-TT000
AM833631 OV4-TT000
A0385232 RPTC11-14
A0314476 RPTC11-10
AL040883 DKFZP434H
A0142135 HS_3157.B
A0215052 HS_3744.A
A2337152 LM0067B09
BI210793 EST528833
A2476585 LM0295H03
A2325063 LM0047J15
A0602696 HS_2116.A
A0306368 HS_2050.A
AM038199 EST279836
A0389869 RPTC11-14
A2289880 RPTC1-23-1
A2343568 LM0064F02
BI093008 602858287
AL439739 T3 end of
AL440744 T7 end of
A0806020 HS_3337.A
A0403329 HS_5061.B
A2385697 LM0144M15
AV006577 AV006577
A2456884 LM0260B02
D12416 HUM007M139
A0682180 HS_5503.B
AA920661 VY48C01.F
BE060601 F125D01.X
AA398717 zt75307.s
R933225 yq11a12.s1
H02904 yj39f03.s1
AA448276 zW83301.S
AM202805 IJ2312.Y
AA184553 mLS3B06.F
A0401672 HS_5044.B
A0977126 RPTC1-23-4
A2072579 RPTC1-23-4
BE779264 602665946
A2650467 LM0520J07
BI222905 60294167
BE394042 601312405
BE613352 601452454
BE548526 601075921
BE957936 601654119
BE966737 602286714
BG617252 602615468
AV357339 AV357339
A2481861 LM0306N23
A2492553 LM0326B17
N81375 T8EST245H1
A0070825 HS_2256.A
Z14476 CEL132 Chr
AI938247 sc41901.Y
AU022999 AU022999
AA787570 n8B04A1.r
BE068478 H3066A04-
BE081507 H3066A04-

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-08-821-324-7

Query Match
Best Local Similarity 78.3%; DB 4; Length 530;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtcacagaatta 25
||||||| ||| ||||| |||
DB 414 CATTGTCGTTAGTTTCAGATTTA 392

RESULT 44
US-09-295-028-7/c
Sequence 7, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 530
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-7

Query Match
Best Local Similarity 60.0%; Score 15; DB 4; Length 530;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtcacagaatta 25
||||||| ||| ||||| |||
DB 414 CATTGTCGTTAGTTTCAGATTTA 392

RESULT 45
US-09-106-582-7/c
Sequence 7, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
THERAPY
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-09-106-582-7

Query Match
Best Local Similarity 60.0%; Score 15; DB 4; Length 530;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtcacagaatta 25
||||||| ||| ||||| |||
DB 414 CATTGTCGTTAGTTTCAGATTTA 392

Search completed: December 26, 2001, 12:20:06
Job time: 6963 sec

TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: ICS-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-3415
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 6525 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 573..5684
US-08-762-500-74

Query Match 61.6%; Score 15.4; DB 3; Length 6525;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgacgaagcacagaatta 25
DB 4890 GCCTTGTGCGGGGTCAACATCA 4914

RESULT 42
US-08-975-762-7/c
Sequence 7, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-08-975-762-7

Query Match 60.0%; Score 15; DB 4; Length 530;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 catctgcgaagcacagaatta 25
DB 414 CATTCGCTTAGTTTCAGATTTA 392

RESULT 43
US-08-821-324-7/c
Sequence 7, Application US/08821324
Patent No. 6231869
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

Query Match 61.6%; Score 15.4; DB 2; Length 2352;
Best Local Similarity 76.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgtcgaagtcacagaatla 25
||| ||||| ||||| |||
DB 2351 GACATTATCGCAGCTCTCAGATTTA 2327

RESULT 39

US-08-665-259-24
; Sequence 24, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..5053
; US-08-665-259-24

Query Match 61.6%; Score 15.4; DB 3; Length 5894;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgtcgaagtcacagaatla 25
||| ||||| ||||| |||
DB 4259 GCCTTTCGCGGGGTCCACAGATCA 4283

RESULT 40
US-08-762-500-24
; Sequence 24, Application US/08762500
; Patent No. 6030806

; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PC7/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..5053
; US-08-762-500-24

Query Match 61.6%; Score 15.4; DB 3; Length 5894;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgtcgaagtcacagaatla 25
||| ||||| ||||| |||
DB 4259 GCCTTTCGCGGGGTCCACAGATCA 4283

RESULT 41

US-08-762-500-74
; Sequence 74, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

COMPUTER: PC compatible
OPERATING SYSTEM: DOS 6.2
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252,493C
FILING DATE: June 1, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Fidel, Seth A.
REGISTRATION NUMBER: 38,449
REFERENCE/DOCKET NUMBER: ALX-138
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 776-1790
TELEFAX: (203) 772-3655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1592 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: Porcine E-selectin
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-252-493C-1

Query Match 61.6%; Score 15.4; DB 2; Length 1592;
Best Local Similarity 76.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 gacattgtcgaagtcacagaatta 25
||| |||| || |||| |||| ||||
Db 860 GACGGTGCACATGTCACACGCTTTA 836

RESULT 37
US-09-276-197-1/c
Sequence 1, Application US/09276197
Patent No. 6040428
GENERAL INFORMATION:
APPLICANT: Rollins, Scott
APPLICANT: Rother, Russell P.
APPLICANT: Evans, Mark J.
APPLICANT: Matis, Louis A.
TITLE OF INVENTION: PORCINE E-SELECTIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seth A. Fidel
STREET: 25 Science Park, Box 15
CITY: New Haven
STATE: Connecticut
COUNTRY: USA
ZIP: 06511
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 750 Kb storage
COMPUTER: PC compatible
OPERATING SYSTEM: DOS 6.2
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,197
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/252,493
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fidel, Seth A.
REGISTRATION NUMBER: 38,449
REFERENCE/DOCKET NUMBER: ALX-138

TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 776-1790
TELEFAX: (203) 772-3655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1592 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: Porcine E-selectin
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-276-197-1

Query Match 61.6%; Score 15.4; DB 3; Length 1592;
Best Local Similarity 76.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 gacattgtcgaagtcacagaatta 25
||| |||| || |||| |||| ||||
Db 860 GACGGTGCACATGTCACACGCTTTA 836

RESULT 38
US-08-889-909A-21/c
Sequence 21, Application US/0889909A
Patent No. 5853977
GENERAL INFORMATION:
APPLICANT: Dalle, Barbara
APPLICANT: Fan, Xuedong
APPLICANT: Lundell, Daniel
APPLICANT: Lunn, Charles
APPLICANT: Tan, Jimmy
APPLICANT: Zavodny, Paul
TITLE OF INVENTION: Mammalian TNF- α Convertases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering-Plough Corporation
STREET: 2000 Galloping Hill Road
CITY: Kenilworth
STATE: New Jersey
COUNTRY: USA
ZIP: 07033-0530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/889,909A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021,710
FILING DATE: 12-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dulak, No. 5853977man C.
REGISTRATION NUMBER: 31,608
REFERENCE/DOCKET NUMBER: JB0601
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908 298 2906
TELEFAX: 908 298 5388
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2352 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-889-909A-21

STRANDEDNESS: double
TOPOLOGY: linear
US-09-340-812-5

Query Match 62.4%; Score 15.6; DB 3; Length 3360;
Best Local Similarity 81.8%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaa 22
|||||
DB 247 GACATTGTCGATTTCACAAA 268

RESULT 34
US-08-916-443A-15
Sequence 15, Application US/08916443A
Patent No. 6001986
GENERAL INFORMATION:
APPLICANT: Yong Sig KIM
APPLICANT: Sun Chung PARK
APPLICANT: Soo Kyung OH
APPLICANT: Hosull LEE
APPLICANT: Jeong Woo CHO
APPLICANT: Chang H. CHUNG
TITLE OF INVENTION: Antiviral proteins, Amarandin 1 and 2, from
TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,443A
FILING DATE: 22 AUG 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 1942/18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 nucleic acids
TYPE: nucleotides
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-916-443A-15

Query Match 61.5%; Score 15.4; DB 3; Length 946;
Best Local Similarity 76.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaa 25
|||||
DB 696 GAAATGTTTCAAGTCGAGAAATTA 720

RESULT 35
US-08-916-443A-16
Sequence 16, Application US/08916443A

Patent No. 6001986
GENERAL INFORMATION:
APPLICANT: Yong Sig KIM
APPLICANT: Sun Chung PARK
APPLICANT: Soo Kyung OH
APPLICANT: Hosull LEE
APPLICANT: Jeong Woo CHO
APPLICANT: Chang H. CHUNG
TITLE OF INVENTION: Antiviral proteins, Amarandin 1 and 2, from
TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,443A
FILING DATE: 22 AUG 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 1942/18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1116 nucleic acids
TYPE: nucleotides
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-916-443A-16

Query Match 61.6%; Score 15.4; DB 3; Length 1116;
Best Local Similarity 76.0%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaa 25
|||||
DB 855 GAAATGTTTCAAGTCGAGAAATTA 879

RESULT 36
US-08-252-493C-1/c
Sequence 1, Application US/08252493C
Patent No. 5891645
GENERAL INFORMATION:
APPLICANT: Rollins, Scott
APPLICANT: Rother, Russell P.
APPLICANT: Evans, Mark J.
APPLICANT: Matlis, Louis A.
TITLE OF INVENTION: PORCINE E-SELECTIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seth A. Fidel
STREET: 25 Science Park, Box 15
CITY: New Haven
STATE: Connecticut
COUNTRY: USA
ZIP: 06511
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 750 Kb storage

US-09-232-200-78

Query Match 62.4%; Score 15.6; DB 4; Length 1953;
Best Local Similarity 81.8%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 catgtgcgaagtcacagaatt 24
||||| | | | | | | | | |
Db 1912 CATTGTCGACGACCAAAATT 1891

RESULT 31

US-09-232-197-78/C
; Sequence 78, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-2193MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 78
; LENGTH: 1953
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-232-197-78

Query Match 62.4%; Score 15.6; DB 4; Length 1953;
Best Local Similarity 81.8%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 catgtgcgaagtcacagaatt 24
||||| | | | | | | | | |
Db 1912 CATTGTCGACGACCAAAATT 1891

RESULT 32

US-08-938-546-5
; Sequence 5, Application US/08938546
; Patent No. 6004556
; GENERAL INFORMATION:
; APPLICANT: Fosberry, Andrew P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Nicholas, Richard O.
; TITLE OF INVENTION: NO. 6004556el rsbu-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/938,546
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:

INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-938-546-5

Query Match 62.4%; Score 15.6; DB 3; Length 3360;
Best Local Similarity 81.8%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgacatttcacaaaa 22
||||| | | | | | | | | |
Db 247 GACATTGTCGATTTACAAAA 268

RESULT 33

US-09-340-812-5
; Sequence 5, Application US/09340812
; Patent No. 6111078
; GENERAL INFORMATION:
; APPLICANT: Fosberry, Andrew P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Nicholas, Richard O.
; TITLE OF INVENTION: NO. 6111078el rsbu-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/340,812
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/938,546
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3360 base pairs
; TYPE: nucleic acid

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: P50551
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1020 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-938-546-3

Query Match 62.4%; Score 15.6; DB 3; Length 1020;
Best Local Similarity 81.8%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaa 22
||||| 11111111 11111111
Db 148 GACATTGTGCATATTCACAAA 169

RESULT 28
US-09-340-812-3
Sequence 3, Application US/09340812
Patent No. 6111078
GENERAL INFORMATION:
APPLICANT: Fosberry, Andrew P.
APPLICANT: Lawlor, Elizabeth J.
APPLICANT: Nicholas, Richard O.
TITLE OF INVENTION: No. 6111078el rsbu-1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/340,812
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/938,546
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: P50551
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1020 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-340-812-3

Query Match 62.4%; Score 15.6; DB 3; Length 1020;
Best Local Similarity 81.8%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaa 22
||||| 11111111 11111111
Db 148 GACATTGTGCATATTCACAAA 169

RESULT 29
US-09-232-191-28/c
Sequence 28, Application US/09232191
Patent No. 6284487
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WH197-21p3ME
CURRENT APPLICATION NUMBER: US/09/232,191
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 28
LENGTH: 1953
TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-09-232-191-28

Query Match 62.4%; Score 15.6; DB 4; Length 1953;
Best Local Similarity 81.8%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 catgttcgaagtcacagaatt 24
||||| 11111111 11111111
Db 1912 CATGTTCGACGACACAAAATT 1891

RESULT 30
US-09-232-200-78/c
Sequence 78, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21p3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 78
LENGTH: 1953
TYPE: DNA
ORGANISM: Caenorhabditis elegans

100

Query Match	63.28;	Score 15.8;	DB 1;	Length 3220;
Best Local Similarity	89.58;	Pred. No. 70;		

DB 3110 ACAGTGTCTCAGTCCAGATATA 3087

RESULT 17

US-09-381-862-5/C
; Sequence 5, Application US/09381862
; Patent No. 6245906
; GENERAL INFORMATION:
; APPLICANT: Ueyama, Hiroshi
; APPLICANT: Abe, Kanako
; APPLICANT: Keshi, Hiroyuki
; APPLICANT: Matsubisa, Akio
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS
; TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gersteln, Murray & Borum
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/381,862
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1997-71077
; FILING DATE: 25-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP98/01288
; FILING DATE: 23-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Cawley, Jr., Thomas A.
; REGISTRATION NUMBER: 40,944
; REFERENCE/DOCKET NUMBER: 19036/36274
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; STRAIN: Clinical Isolate Sp-26-46
; US-09-381-862-5

Query Match 64.0%; Score 16; DB 4; Length 6688;
Best Local Similarity 79.2%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gacatgtcgcaagtcacgaatt 24
DB 4228 GACGTGTGCAATTCACGTAATT 4205

RESULT 18
US-08-781-562-2/C
; Sequence 2, Application US/08781562
; Patent No. 5783589
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,562
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0181 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
; US-08-781-562-2

Query Match 63.2%; Score 15.8; DB 1; Length 1188;
Best Local Similarity 89.5%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acatgtcgcaagtcacag 20
DB 353 ACATGTCTCAGTCCAGTACAG 335

RESULT 19
US-08-687-379-1
; Sequence 1, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Mitsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Steirer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 kb storage

APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ. ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: brain
FEATURE:
NAME/KEY: CDS
LOCATION: 73..2736
US-08-687-379-11

Query Match 64.8%; Score 16.2; DB 1; Length 2989;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 catgtcgcaagtcacagaat 23
|||||
DB 2939 CATGTGCGAGTCACAGACT 2919

RESULT 15

US-08-617-860B-3/c
Sequence 3, Application US/08617860B
Patent No. 6133506
GENERAL INFORMATION:
APPLICANT: Tyfeler, R., Bautor, J., Bothmann, H., Flisak, E.,
APPLICANT: Hyricke-Grandpierre, C., Klein, B., Martini, N.,
APPLICANT: Moller, A., Schulte, W., Voeltz, M., Walek, J.,
APPLICANT: Schell, J.
TITLE OF INVENTION: Promoters
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
STREET: 1140 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,860B
FILING DATE: 01-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02950
FILING DATE: 05-SEP-1994
APPLICATION NUMBER: DE P4329951.2
FILING DATE: 04-SEP-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3350 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Brassica napus
IMMEDIATE SOURCE:
LIBRARY: genomic lambda FIX II

CLONE: BnACCseg10
FEATURE:
NAME/KEY: Startcodon
LOCATION: 2611..2613
FEATURE:
NAME/KEY: CDS
LOCATION: join(2611..2908, 3001..3341)
US-08-617-860B-3

Query Match 64.0%; Score 16; DB 3; Length 3350;
Best Local Similarity 79.2%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaat 24
|||||
DB 443 GACCTGTGCGAATGAGAGAAATT 420

RESULT 16

US-09-032-365A-11/c
Sequence 11, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502th, Michael
APPLICANT: Nishina, Patsy
APPLICANT: Nagart, Juergen
APPLICANT: No. 6114502en-Trauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650 327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 5994 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-032-365A-11

Query Match 64.0%; Score 16; DB 3; Length 5994;
Best Local Similarity 79.2%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaat 25

```

; SEQ ID NO 48
; LENGTH: 1771
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1193)..(1387)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1709)..(1771)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L20788 Genbank
; DATABASE ENTRY DATE: 1996-04-18
; US-09-166-203-48

```

```

Query Match      64.8%; Score 16.2; DB 2; Length 1771;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 attgtgcgaagtcacagaatt 24
   ||| ||||| ||||| ||||| |||||
DB 306 agtgaagcaagtcacagaatt 326

```

```

RESULT 12
; US-09-377-309-48
; Sequence 48, Application US/09377309B
; Patent No. 6258790
; GENERAL INFORMATION:
; APPLICANT: Bennett, C. Frank
; APPLICANT: Condon, Tom P.
; APPLICANT: Cowsett, Lex M.
; TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
; FILE REFERENCE: ISPH-0390
; CURRENT APPLICATION NUMBER: US/09/377, 309B
; CURRENT FILING DATE: 1999-08-19
; EARLIER APPLICATION NUMBER: 09/166, 203
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 99
; SEQ ID NO 48
; LENGTH: 1771
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1193)..(1387)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1709)..(1771)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L20788 Genbank
; DATABASE ENTRY DATE: 1996-04-18
; US-09-377-309-48

```

```

Query Match      64.8%; Score 16.2; DB 4; Length 1771;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 attgtgcgaagtcacagaatt 24
   ||| ||||| ||||| ||||| |||||
DB 306 agtgaagcaagtcacagaatt 326

```

```

RESULT 13
; US-08-687-379-9/c
; Sequence 9, Application US/08687379
; Patent No. 575697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ullsch, Andreas
; APPLICANT: Bach, Alfred

```

```

; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687, 379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..2736
; US-08-687-379-9

```

```

Query Match      64.8%; Score 16.2; DB 1; Length 2989;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 catgtgcgaagtcacagaatt 23
   ||||| ||||| ||||| ||||| |||||
DB 2939 CATGTGCGAGTCTCAGAGT 2919

```

```

RESULT 14
; US-08-687-379-11/c
; Sequence 11, Application US/08687379
; Patent No. 575697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ullsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:

```

ATTORNEY/AGENT INFORMATION:
NAME: Cantor, Herbert I.
REGISTRATION NUMBER: 24,392
REFERENCE/DOCKET NUMBER: P-450-23557
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-0400
TELEFAX: 202-835-0605
TELEX: 440706
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gtgcgaagtcacagaat 23
|||||
Db 17 gtgcgaagtcacagaat 1

RESULT 9
US-08-123-761A-2/c
Sequence 2, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
TITLE OF INVENTION: A SELECTABLE MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-123-761A-2

Query Match 66.4%; Score 16.6; DB 1; Length 1374;

Best Local Similarity 82.6%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acatgtcgaagtcacagaatt 24
|||||
Db 1329 AATTTATCGCAGTCAATAATT 1307

RESULT 10
US-08-123-761A-1/c
Sequence 1, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
TITLE OF INVENTION: A SELECTABLE MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5198 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-123-761A-1

Query Match 66.4%; Score 16.6; DB 1; Length 5198;
Best Local Similarity 82.6%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acatgtcgaagtcacagaatt 24
|||||
Db 4891 AATTTATCGCAGTCAATAATT 4869

RESULT 11
US-09-166-203-48
Sequence 48, Application US/09166203A
Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Bennet, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowser, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0323
CURRENT APPLICATION NUMBER: US/09/166,203A
FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60

CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619)535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 71.2%; Score 17.8; DB 1; Length 3083;
Best Local Similarity 90.5%; Pred. No. 7.5;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 catgtcgcaagtcacagaat 23
|||||
Db 3034 CATTGTCGCAAGTCTCAGAGT 3014

RESULT 7
US-08-486-269A-5/c
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
TITLE OF INVENTION: AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 71.2%; Score 17.8; DB 2; Length 3083;
Best Local Similarity 90.5%; Pred. No. 7.5;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 catgtcgcaagtcacagaat 23
|||||
Db 3034 CATTGTCGCAAGTCTCAGAGT 3014

RESULT 8
US-07-944-259-2/c
Sequence 2, Application US/07944259
Patent No. 5885792
GENERAL INFORMATION:
APPLICANT: Ifuku, Ohji
APPLICANT: Haze, Shintaro
APPLICANT: Kishimoto, Jiro
APPLICANT: Nakahama, Kazuo
TITLE OF INVENTION: BIOTIN OPERON
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wegner, Cantor, Mueller & Player
STREET: 1233 20th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-8218
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/944,259
FILING DATE: 19920914
CLASSIFICATION: 435


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=1154
OTHER INFORMATION: /EC_number=2.3.1.47
OTHER INFORMATION: /product="KAPA synthase"
OTHER INFORMATION: /evidence="EXPERIMENTAL"
OTHER INFORMATION: /gene="bioF"
OTHER INFORMATION: /number=2
FEATURE:
NAME/KEY: OTHER INFORMATION: /standard_name="8-Amino-7-oxononanoate synthase"
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3043
OTHER INFORMATION: /EC_number=6.3.3.3
OTHER INFORMATION: /product="DPR synthase"
OTHER INFORMATION: /evidence="EXPERIMENTAL"
OTHER INFORMATION: /gene="bioD"
OTHER INFORMATION: /number=4
FEATURE:
NAME/KEY: OTHER INFORMATION: /standard_name="Dethiobiotin synthase"
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name="bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name="bioD RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match
Best Local Similarity 100.0%; Score 25; DB 3; Length 5872;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgcgaagtcacagaata 25
|||||
Db 137 GACATTGCGCAGTCACAGATTA 161
```

```

GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 1; Length 3083;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 callytcgcaagtcacagaat 23
|||||
Db 3034 CATTGTCGCAAGTCACAGAT 3014

RESULT 6
US-08-481-206-5/c
Sequence 5; Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "Involved in pimejoyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAFA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: aminoTransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "orf1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-Apr-1993
US-08-411-768B-1

```

Query Match 100.0%; Score 25; DB 3; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0027;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 gacattgtcgaatgcacagaatga 25
DB 137 GACATTGTGCAATGCACAGATTA 161

```

```

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
Version 5.1

```

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 21 GACATTGTGCGCAAGTCACAGAAATTA 45

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 21 GACATTGTGCGCAAGTCACAGAAATTA 45

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; OF Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:24:03 : Search time 141.92 Seconds

(Without alignments)
39.695 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacattgcgaagcagcagacgaattta 25

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.8	71.2	3083	1	US-07-718-575-5
6	17.8	71.2	3083	1	US-08-481-206-5
7	17.8	71.2	3083	2	US-08-486-269A-5
8	17	68.0	17	2	US-07-944-259-2
9	16.6	66.4	1374	1	US-08-123-761A-2
10	16.6	66.4	5198	1	US-08-123-761A-1
11	16.2	64.8	1771	4	US-09-166-203-48
12	16.2	64.8	1771	4	US-09-377-309-48
13	16.2	64.8	2969	1	US-08-687-379-9
14	16.2	64.8	2969	1	US-08-687-379-11
15	16	64.0	3350	3	US-08-617-860B-3
16	16	64.0	5994	3	US-09-332-365A-11
17	16	64.0	6688	4	US-09-381-862-5
18	15.8	63.2	1188	1	US-08-781-562-2
19	15.8	63.2	2946	1	US-08-687-379-1
20	15.8	63.2	2946	1	US-08-687-379-3
21	15.8	63.2	3220	1	US-08-254-573-1
22	15.6	62.4	350	1	US-08-473-040-6
23	15.6	62.4	350	5	PCT-US95-09715-6
24	15.6	62.4	633	4	US-09-328-111-497
25	15.6	62.4	1002	3	US-08-338-546-1
26	15.6	62.4	1002	3	US-09-340-812-1
27	15.6	62.4	1020	3	US-08-938-546-3
28	15.6	62.4	1020	3	US-09-340-812-3
29	15.6	62.4	1953	4	US-09-232-191-28
30	15.6	62.4	1953	4	US-09-232-200-78
31	15.6	62.4	1953	4	US-09-232-197-78
32	15.6	62.4	3360	3	US-08-938-546-5
33	15.6	62.4	3360	3	US-09-340-812-5
34	15.4	61.6	3460	3	US-08-916-443A-15
35	15.4	61.6	1116	3	US-08-916-443A-16
36	15.4	61.6	1592	2	US-08-252-493C-1
37	15.4	61.6	1592	2	US-09-276-197-1
38	15.4	61.6	2352	2	US-08-889-909A-21
39	15.4	61.6	5894	3	US-08-665-259-24
40	15.4	61.6	5894	3	US-08-762-500-24
41	15.4	61.6	6525	3	US-08-762-500-74
42	15	60.0	530	4	US-08-975-762-7
43	15	60.0	530	4	US-08-821-334-7
44	15	60.0	530	4	US-09-295-028-7
45	15	60.0	530	4	US-09-106-582-7
46	15	60.0	1749	1	US-08-116-098-1
47	15	60.0	6012	1	US-08-375-709-14
48	15	60.0	6012	1	US-08-752-929-14
49	15	60.0	8010	4	US-09-521-526-2
50	15	60.0	8010	5	PCT-US95-11859-2
51	15	60.0	37895	1	US-08-375-709-1
52	15	60.0	37895	1	US-08-752-929-1
53	15	60.0	37895	3	US-09-090-793-1
54	15	60.0	56516	2	US-08-996-306-1
55	15	60.0	56516	4	US-09-338-907-1
56	15	60.0	56516	4	US-09-338-907-179
57	14.8	58.4	246240	2	US-08-724-394A-20
58	14.8	58.4	246240	2	US-08-724-394A-21
59	14.8	58.4	246240	2	US-08-724-394A-22
60	14.6	58.4	450	1	US-08-374-843B-4
61	14.6	58.4	450	1	US-08-374-843B-18
62	14.6	58.4	450	2	US-08-905-420-4
63	14.6	58.4	450	2	US-08-905-420-18
64	14.6	58.4	720	6	US-08-374-843B-9
65	14.6	58.4	992	1	US-08-905-420-9
66	14.6	58.4	992	2	US-08-905-420-9
67	14.6	58.4	992	1	US-08-374-843B-16
68	14.6	58.4	1758	2	US-08-905-420-16
69	14.6	58.4	1758	3	US-08-836-567-3
70	14.6	58.4	2026	2	US-08-993-228-3
71	14.6	58.4	2181	6	US-08-836-567-9
72	14.6	58.4	2360	3	US-08-836-567-9
73	14.6	58.4	2680	1	US-08-588-821-1
74	14.6	58.4	2680	1	US-08-915-214-1
75	14.6	58.4	2680	2	US-09-005-532-1
76	14.6	58.4	2680	3	US-09-255-888-1
77	14.6	58.4	2931	4	US-09-056-105-15
78	14.6	58.4	3348	1	US-08-222-616-35
79	14.6	58.4	3348	5	PCT-US95-04228-35
80	14.4	57.6	195	4	US-08-945-625A-1
81	14.4	57.6	491	4	US-09-400-208B-23
82	14.4	57.6	1248	4	US-09-198-956-5
83	14.4	57.6	1248	4	US-09-198-956-5
84	14.4	57.6	1272	2	US-08-972-258-3
85	14.4	57.6	1272	4	US-08-972-258-3
86	14.4	57.6	1380	1	US-08-484-324-1
87	14.4	57.6	1380	1	US-08-486-409-1
88	14.4	57.6	1380	2	US-08-825-784-1
89	14.4	57.6	1380	5	PCT-US92-01899-1
90	14.4	57.6	1426	3	US-09-121-425-3
91	14.4	57.6	1430	4	US-08-855-910-10
92	14.4	57.6	1484	4	US-09-402-002-4
93	14.4	57.6	1690	4	US-08-943-731-166
94	14.4	57.6	2051	2	US-08-691-814B-11
95	14.4	57.6	2075	1	US-08-167-628-1
96	14.4	57.6	2075	1	US-08-386-660-1
97	14.4	57.6	2075	1	US-08-459-717-1
98	14.4	57.6	2075	1	US-08-712-302-1
99	14.4	57.6	2075	2	US-08-880-031-1
100	14.4	57.6	2075	3	US-09-097-179-1

us-09-396-196f-1.rng

Sequence 470 BP; 129 A; 108 C; 117 G; 116 T; 0 other;

Sequence 470 BP; 129 A; 108 C; 117 G; 116 T; 0 other;

Matches	19;	Conservative	0;	Mismatches	5;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

```

QY      1  gacatgtgtcgaagtcacagaatt 24
        |  || ||||  |||||  ||||
Db      18  gtcaagtgtcttaagtcactgaatt 41

```

Search completed: December 26, 2001, 12:17:28
Job time: 6809 sec

PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 PS
 PS Claim 25; SEQ ID No 159; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 470 BP; 129 A; 108 C; 117 G; 116 T; 0 other;
 XX
 Query Match 64.0%; Score 16; DB 22; Length 470;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 gacattgtcgcagtcacagaatt 24
 1 11 1111 111111 11111
 DB 18 gtcaagtgtcctaagtcactgaatt 41
 XX
 RESULT 44
 AA131474
 ID AA131474 standard; DNA; 470 BP.
 XX
 AC AA131474;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #160 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-48897/53.
 XX

PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 160; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 470 BP; 129 A; 108 C; 117 G; 116 T; 0 other;
 XX
 Query Match 64.0%; Score 16; DB 22; Length 470;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 gacattgtcgcagtcacagaatt 24
 1 11 1111 111111 11111
 DB 18 gtcaagtgtcctaagtcactgaatt 41
 XX
 RESULT 45
 AA100164
 ID AA100164 standard; DNA; 470 BP.
 XX
 AC AA100164;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Probe #155 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 PS
 PS Claim 25; SEQ ID No 155; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.


```

XX Homo sapiens.
OS
XX WO200009525-A2.
XX
XX 24-FEB-2000.
XX
XX 03-AUG-1999; 99WO-US17712.
XX
XX 03-AUG-1998; 98US-0095212.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX Nyce JW:
XX
XX WPI: 2000-205971/18.
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
XX Disclosure: Page 1221-1231; 1343pp; English.
XX
XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antisthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA2313 to AAA3312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA2333 to AAA3392) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
XX Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other:
SQ

```

Query Match 64.8%; Score 16.2; DB 21; Length 40298;
 Best Local Similarity 85.7%; Pred. No. 3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 5 ttgtcgcaatcacagaata 25
    ||| | ||||| ||||| |||
DB 11514 ttgcccaaatcacagaata 11534

```

RESULT 40
 AAH30726/C
 ID AAH30726 standard; cDNA; 395 BP.
 XX
 XX AAH30726;
 XX
 XX 27-JUL-2001 (first entry)
 XX
 XX Human colon cancer cell line Km12L4-A cDNA library derived sequence #660.
 XX
 XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
 KW

```

KW detection: colon cancer cell line Km12L4-A; ss.
XX
XX Homo sapiens.
OS
XX WO200018916-A2.
XX
XX 06-APR-2000.
XX
XX 23-SEP-1999; 99WO-US22226.
XX
XX 28-SEP-1998; 98US-0102161.
XX
XX 28-SEP-1998; 98US-0102180.
XX
XX 29-SEP-1998; 98US-0102380.
XX
XX 08-OCT-1998; 98US-0103815.
XX
XX 27-OCT-1998; 98US-0105877.
XX
XX (CHIR) CHIRON CORP.
XX
XX (HSE-) HSECO INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX Ianson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI: 2000-293155/25.
XX
XX Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
XX
XX Claim 1; Page 374; 502pp; English.
XX
XX The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Km12L4-A cDNA library.
XX
XX Sequence 395 BP; 124 A; 63 C; 90 G; 106 T; 12 other:
SQ

```

Query Match 64.0%; Score 16; DB 21; Length 395;
 Best Local Similarity 79.2%; Pred. No. 2.1e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY 2 acattgtcgcaagtcacagaata 25
    ||||| |||| | ||||| |
DB 295 ACATTGTACCAATATCCAGAAATTA 272

```

RESULT 41
 AAT67395
 ID AAT67395 standard; DNA; 420 BP.
 XX
 XX AAT67395;
 XX
 XX 08-JUL-1997 (first entry)
 XX
 XX H. pylori surface or membrane protein ORF 14726542.aa.
 XX
 XX Cytoplasmic; vaccine; prevention; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW

PT cancers -
 XX
 PS Disclosure: Page 1212-1219; 1343pp; English.
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytosolic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA33323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.
 XX
 S0 Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;
 XX
 Query Match 64.8%; Score 16.2; DB 21; Length 32351;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 ttgtgcagatcacagaatca 25
 ||| | |||||
 Db 10689 ttgccaaagtcacagaatca 10709
 RESULT 38
 AAF21311
 ID AAF21311 standard; DNA: 40298 BP.
 XX
 AC AAF21311;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2878.
 XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytosolic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 KW
 XX Homo sapiens.
 OS
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX

PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX
 PI NYCE JW;
 XX
 DR WPI: 2000-679539/66.
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure: Page 1305-1315; 1592pp; English.
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytosolic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 CC
 S0 Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;
 XX
 Query Match 64.8%; Score 16.2; DB 21; Length 40298;
 Best Local Similarity 85.7%; Pred. No. 3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 ttgtgcagatcacagaatca 25
 ||| | |||||
 Db 11514 ttgccaaagtcacagaatca 11534
 RESULT 39
 AAA35189
 ID AAA35189 standard; DNA: 40298 BP.
 XX
 AC AAA35189;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.
 XX
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;
 KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

CC The GluR3A receptor DNA is contained within plasmid pBS/humGluR3A
 CC (ATCC 75218) and may be expressed by a mammal recombinant cell. The
 CC DNA can be used to obtain products for evaluating the human nervous
 CC system and particularly for assessing potentially therapeutic
 CC interactions between the AMPA-binding human excitatory amino acid
 CC receptors and natural and synthetic ligands. It can also be used
 CC for detecting the presence and/or location of human GluR receptors
 CC e.g. in brain tissue. The DNA was isolated from human brain CDNA
 CC libraries using rat GluR receptor gene probes.
 XX Sequence 3070 BP; 899 A; 652 C; 738 G; 780 T; 1 other:
 SQ

Query Match 64.8%; Score 16.2; DB 14; Length 3070;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 catgtcgaagtcacagaat 23
 ||||| ||||| ||||| |||||
 Db 2946 CATGTGCGAGTCTCAGACT 2926

RESULT 36
 AAF21307 standard; DNA: 32351 BP.
 ID AAF21307:
 AC AAF21307:
 XX 14-MAR-2001 (first entry)
 DT
 XX Human low adenosine antisense oligonucleotide related sequence #2874.
 DE
 XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 PN
 XX WO200062736-A2.
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000MO-0508020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX
 PI NYCE JW;
 XX
 DR WPI; 2000-679539/66.
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure; Page 1295-1303; 1592pp; English.
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and or activity of target polypeptides associated with

CC Lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (Cf), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other:
 S0

Query Match 64.8%; Score 16.2; DB 21; Length 32351;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ttgtcgaagtcacagaatta 25
 ||| ||||| ||||| |||||
 Db 10689 ttgcccaagtcacagaatta 10709

RESULT 37
 AAA35185 standard; DNA: 32351 BP.
 ID AAA35185:
 AC AAA35185:
 XX 28-JUL-2000 (first entry)
 DT
 XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
 DE
 XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 XX
 OS Homo sapiens.
 PN
 XX WO200009525-A2.
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA NYCE JW;
 XX
 PI NYCE JW;
 XX
 DR WPI; 2000-205971/18.
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or

CC AAT02795-802 are DNA sequences encoding variants of the human glutamate
 CC receptor (GluR) subunits. cDNAs encoding GluR subunits A, B, C, and D,
 CC were isolated from the brain. There are 2 splice variants, "flip" and
 CC "flop", for each subunit that are known. It was found that for GluRA,
 CC GluRB, GluRC and GluRD, the first amino acid of the "flip/flop"-region,
 CC can be glycine (G) or arginine (R), through different RNA splicing.
 CC For the GluRB subunit, it was found that it is possible to have a
 CC glutamine (Q) or an R in the transmembrane domain. Cells transfected with
 CC the DNA are useful for identifying functional ligands for GluR.
 CC Glutamate is an important excitatory neurotransmitter in the central
 CC nervous system and is involved in the pathophysiology of e.g., epilepsy,
 CC schizophrenia and ischemia. GluR ligands have the potential to be
 CC used in the treatment of these conditions.
 CC
 CC Sequence 2989 BP; 868 A; 641 C; 728 G; 752 T; 0 other;

QY 3 catgtcgcaagtcacagat 23
 Db 2939 CATTGTCGAGCTCAGAGT 2919

RESULT 34
 AAT02800/c
 ID AAT02800 standard: cDNA to mRNA; 2989 BP.
 AC AAT02800;
 XX
 DT 26-APR-1996 (first entry)
 XX
 DE Human glutamate receptor subunit cDNA (GluRC/flop).
 XX
 KW glutamate receptor subunit; human; flip: flop: splice variant;
 KM identify ligand; treatment; ischemia; schizophrenia; epilepsy; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 73..2739
 FT CDS /*tag= a
 FT /note= "glutamate receptor subunit"
 XX
 PN W09521188-A1.
 XX
 PD 10-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-EP00290.
 XX
 PR 07-FEB-1994; 94DE-4403666.
 XX
 PA (BADI) BASF AG.
 XX
 PI Bach A, Hoeger T, Lemaire H, Steiner S, Ullrich A;
 XX
 DR WPI: 1995-283726/37.
 DR P-PSDB: AAR84917.
 XX
 PT DNA sequences coding for glutamate receptor sub-unit variants -
 PT useful for the identification of functional ligands for these
 PT receptors and for preparing anti-sense reagents
 XX
 PS Claim 1: Page 54-58; 71pp: English.
 XX
 CC AAT02795-802 are DNA sequences encoding variants of the human glutamate
 CC receptor (GluR) subunits. cDNAs encoding GluR subunits A, B, C, and D,
 CC were isolated from the brain. There are 2 splice variants, "flip" and
 CC "flop", for each subunit that are known. It was found that for GluRA,
 CC GluRB, GluRC and GluRD, the first amino acid of the "flip/flop"-region,
 CC can be glycine (G) or arginine (R), through different RNA splicing.
 CC For the GluRB subunit, it was found that it is possible to have a
 CC glutamine (Q) or an R in the transmembrane domain. Cells transfected with
 CC the DNA are useful for identifying functional ligands for GluR.
 CC Glutamate is an important excitatory neurotransmitter in the central
 CC nervous system and is involved in the pathophysiology of e.g., epilepsy,
 CC schizophrenia and ischemia. GluR ligands have the potential to be
 CC used in the treatment of these conditions.

CC can be glycine (G) or arginine (R), through different RNA splicing.
 CC For the GluRB subunit, it was found that it is possible to have a
 CC glutamine (Q) or an R in the transmembrane domain, before the flip/flop
 CC region, giving a possible 8 conformations. The GluRflopG cDNA
 CC possesses a base change at position 2377, a glycine codon (GGA) is
 CC converted to an arginine codon (AGA) (sic). Cells transfected with the
 CC DNA are useful for identifying functional ligands for GluR. Glutamate
 CC is an important excitatory neurotransmitter in the central nervous
 CC system and is involved in the pathophysiology of e.g., epilepsy,
 CC schizophrenia and ischemia. GluR ligands have the potential to be
 CC used in the treatment of these conditions.
 CC
 CC Sequence 2989 BP; 866 A; 641 C; 730 G; 752 T; 0 other;

QY 3 catgtcgcaagtcacagat 23
 Db 2939 CATTGTCGAGCTCAGAGT 2919

RESULT 35
 AA062694/c
 ID AA062694 standard: cDNA; 3070 BP.
 AC AA062694;
 XX
 DT 29-JUN-1994 (first entry)
 XX
 DE Human GluR3B receptor DNA.
 XX
 KW GluR3B receptor; glutamate receptor; AMPA-binding receptor;
 KM CNS; central nervous system; plasmid pBS/hungluR3B; ds.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 79..2745
 FT CDS /*tag= a
 FT sig_peptide 79..144
 FT /*tag= b
 FT mat_peptide 145..2745
 FT /*tag= c
 XX
 PN EP574257-A.
 XX
 PD 15-DEC-1993.
 XX
 PF 10-JUN-1993; 93EP-0304500.
 XX
 PR 10-JUN-1992; 92US-0896437.
 PR 10-JUN-1992; 92US-0896611.
 PR 10-JUN-1992; 92US-0896612.
 XX
 PA (ALX) ALLELIX BIOPHARMACEUTICALS.
 PA (ELLI/) ELLIOTT C.
 PA (KAMB/) KAMBOU R.
 PA (NUTT/) NUTT S L.
 XX
 PI Elliott C, Kamboj R, Nutt SL;
 XX
 DR WPI: 1993-396803/50.
 DR P-PSDB: AAR45143.
 XX
 PT Polynucleotide(s) encoding AMPA-binding human glutamate receptors
 PT - used for providing prods. for use in screening assays or in
 PT detection and/or localisation
 XX
 PS Claim 1: Page 35-39; 101pp: English.
 XX

XX WPI: 1999-373465/32.
 DR P-PSDB: AAY08966.
 XX Ashbya gossypii purine biosynthesis enzymes and related genes for
 PT production of riboflavin
 PS Claim 20; Page 35-38; 48pp; German.
 CC This invention describes novel enzymes involved in purine biosynthesis
 CC in Ashbya gossypii. The enzymes described in the invention include
 CC phosphoribosylpyrophosphate (PRPP) synthetases KPR1 and KPR2
 CC glutamine-phosphoribosylpyrophosphate amidotransferase (ADE4)
 CC IMP-dehydrogenase (GUA1) and a GMP-synthetase (GUA2). The nucleic acid
 CC sequences encoding the enzymes involved in purine biosynthesis are used
 CC to genetically modify microorganisms, for the production of riboflavin.
 CC Riboflavin, also known as vitamin B2, is essential in humans and animals
 CC and is used to treat inflammation of the mouth and throat mucosal layer
 CC as well as inflammation in skin disorders. This sequence represents the
 CC GUA2 gene which encodes a guanosine-monophosphate synthetase.
 XX
 SQ Sequence 2697 BP; 656 A; 631 C; 788 G; 622 T; 0 other;

Query Match 64.8%; Score 16.2; DB 20; Length 2697;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaa 22
 1 11111111111111111111
 DB 69 AATGCTGCAAGTCACAGAA 49

RESULT 32
 AAT67318/C
 ID AAT67318 standard; cDNA; 2879 BP.
 AC AAT67318;
 XX
 DT 15-DEC-1997 (first entry)
 XX

DE Xenopus beta-signalin 4 cDNA.
 XX
 KW Signalin; Xc-signalin-4; transforming growth factor-beta;
 KW signal transduction; growth; differentiation; pattern formation;
 KW hyperproliferative disorder; tumour suppression; spermatogenesis;
 KW osteogenesis; chondrogenesis; ss.
 XX

OS Xenopus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 258..2045
 FT /*tag= a

WO9722697-A1.

XX 26-JUN-1997.
 XX
 PF 20-DEC-1996; 96WO-US20745.
 XX
 PR 20-DEC-1995; 95US-0580031.
 XX

PA (HARD) HARVARD COLLEGE.
 XX (ONTO-) ONTOCENT INC.
 XX

PI Graf Jm, Jin P, Melton DA, Woolf TM;
 XX

DR WPI: 1997-341688/31.
 DR P-PSDB: AAM18097.
 XX

PT Vertebrate signalin polypeptide and DNA - used to modulate the
 PT growth, differentiation or survival of mammalian cells, e.g. for
 PT treatment of hyperproliferative disorders

PS Claim 40; Page 100-103; 152pp; English.
 XX

CC This cDNA clone codes for Xenopus beta-signalin 4 (AAM18097), a
 CC novel protein that modulates the signal transduction activity of a
 CC transforming growth factor receptor. It is an example of a novel
 CC family of vertebrate genes (see also AAT67315-27) whose products (see
 CC also AAM18094-W18106) have broad apparent involvement in mesoderm
 CC induction, tumour suppression and the formation and maintenance
 CC of ordered spatial arrangements of differentiated tissues. The
 CC clone was isolated from Xenopus embryos by PCR amplification (see
 CC AAT67328-29). Isolated nucleic acids can be used to express signalin
 CC polypeptides in prokaryotic or eukaryotic hosts, or in transgenic
 CC animals, or can be delivered to a cell for use in gene therapy or
 CC antisense therapy. Mutation or misexpression of a signalin gene
 CC can be detected using probes, PCR primers or LCR primers based on
 CC signalin genes. This forms the basis of a claimed diagnostic assay
 CC to identify cells at risk of a disorder characterised by unwanted
 CC cell proliferation or differentiation.
 XX

SQ Sequence 2879 BP; 693 A; 759 C; 721 G; 706 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 2879;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacaga 21
 11111111111111111111
 DB 1716 GACACTGTGCGCAAGTCACAACA 1696

RESULT 33
 AAT02799/C
 ID AAT02799 standard; cDNA to mRNA; 2989 BP.
 AC AAT02799;
 XX
 DT 26-APR-1996 (first entry)
 XX

DE Human glutamate receptor subunit cDNA (GluRC11pc).
 XX

KW glutamate receptor subunit; human; fltp; fltp; splice variant;
 KW identify ligand; treatment; ischemia; schizophrenia; epilepsy; ss.
 XX
 OS Homo sapiens.
 XX

FH Key Location/Qualifiers
 FT CDS 73..2739
 FT /*tag= a
 FT /note= "glutamate receptor subunit"

WO9521188-A1.

XX 10-AUG-1995.
 XX
 PD 27-JAN-1995; 95WO-EP00290.
 XX
 PF 07-FEB-1994; 94DE-4403666.
 XX
 PR (BADI) BASF AG.
 XX

PA Bach A, Hoeger T, Lemaire H, Sterrer S, Ullsch A;
 XX

PI WPI: 1995-283726/37.
 DR P-PSDB: AAR84916.
 XX

PT DNA sequences coding for glutamate receptor sub-unit variants -
 PT useful for the identification of functional ligands for these
 PT receptors and for preparing anti-sense reagents

PS Claim 1; Page 46-50; 71pp; English.
 XX

```

FT      /tag= c
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 543..545
FT      /tag= d
FT      /transl_except= CGG encodes Lys.
FT      misc_difference 663..665
FT      /tag= e
FT      /transl_except= CGT encodes Lys.
FT      misc_difference 819..821
FT      /tag= f
FT      /transl_except= CGG encodes Lys.
FT      misc_difference 915..917
FT      /tag= g
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 963..965
FT      /tag= h
FT      /transl_except= CGG encodes Lys.
FT      misc_difference 996..998
FT      /tag= i
FT      /transl_except= AGG encodes Lys.
FT      misc_difference 1203..1205
FT      /tag= j
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 1230..1232
FT      /tag= k
FT      /transl_except= CGA encodes Lys.
FT      misc_difference 1431..1433
FT      /tag= l
FT      /transl_except= CGT encodes Lys.
FT      misc_difference 1443..1445
FT      /tag= m
FT      /transl_except= CGT encodes Lys.
FT      misc_difference 1484..1486
FT      /tag= n
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 1533..1535
FT      /tag= o
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 1623..1625
FT      /tag= p
FT      /transl_except= CGA encodes Lys.
FT      misc_difference 1704..1706
FT      /tag= q
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 1782..1784
FT      /tag= p
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 1788..1790
FT      /tag= q
FT      /transl_except= CGA encodes Lys.
FT      misc_difference 1872..1874
FT      /tag= r
FT      /transl_except= CGT encodes Lys.
FT      misc_difference 1992..1994
FT      /tag= s
FT      /transl_except= AGG encodes Lys.
FT      misc_difference 2004..2006
FT      /tag= t
FT      /transl_except= CGA encodes Lys.
FT      misc_difference 2091..2093
FT      /tag= u
FT      /transl_except= AGA encodes Lys.
FT      misc_difference 2100..2102
FT      /tag= v
FT      /transl_except= AGA encodes Lys.
XX      W09321326-A.
XX      28-OCT-1993.
XX      13-APR-1993; 93WO-FR00367.
XX      13-APR-1992; 92FR-0004491.
PR

```

```

XX      (ORSA-) ORSAN.
PA
XX      Kazmaier M, Lacroite F, Mignotte-vieux C, Minet M;
PI      Pompon D;
XX      WPI: 1993-351736/44.
DR      P-PSDB; AAR43581.
XX
PT      New DNA encoding plant NADPH cytochrome P450 reductase - cloned
PT      by functional complementation in yeast, also recombinant enzyme
PT      useful in P450 mediated bioconversion processes
XX
PS      Claim 11; Figure 9; 79pp; French.
XX
CC      A new method for determining whether a DNA sequence encodes an NADPH
CC      cytochrome P450 reductase involves transforming yeasts with plasmids
CC      of a total cDNA bank of plant(s). The yeasts used in the procedure
CC      are incapable of producing their own NADPH cytochrome P450
CC      reductase. They are then exposed to a cytochrome P450 inhibitor at a
CC      level which is lethal to the yeast cells but not to cells which,
CC      because of the transformation, now contain an active NADPH
CC      cytochrome P450 reductase. Surviving clones are then isolated and
CC      plasmid DNA extracted. The gene is inserted into the plasmid at a
CC      site which places it under the control of an inducible promoter.
XX
SQ      Sequence 2114 BP; 603 A; 400 C; 542 G; 569 T; 0 other;

Query Match      64.8%; Score 16.2; DB 14; Length 2114;
Best Local Similarity 85.7%; Pred. No. 2,1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 attgtgcgaagtcacagaatt 24
Db      1338 atcgttcgaagtcacagaagaat 1358

RESULT 31
AAx78082/c
ID AAx78082 standard; DNA: 2697 BP.
XX
AC      AAx78082;
XX
DT      19-AUG-1999 (first entry)
XX
DE      A. gossypii GUA2 DNA.
XX
KW      Purine biosynthesis; phosphoribosylpyrophosphate synthetase; KPR1; KPR2;
KW      glutamine-phosphoribosylpyrophosphate amidotransferase; ADB4; GUA1;
KW      IMP-dehydrogenase; GUA1; GMP-synthetase; GUA2; PRPP synthetase; mouth;
KW      riboflavin; vitamin B2; treatment; inflammation; throat; mucosal layer;
KW      skin disorder; ss.
XX
OS      Ashbya gossypii.
XX
FH      Key Location/Qualifiers
FT      CDS 456..2033
FT      /tag= a
FT      /product= "Guanosine-monophosphate synthetase"
XX
XX      EP927761-A2.
XX      07-JUL-1999.
XX
XX      08-DEC-1998; 98EP-0123331.
XX      23-DEC-1997; 97DE-1057755.
XX      (BADI ) BASF AG.
XX
XX      Hoeffken HW, Jimenez A, Pompejus M, Revuelta Doval JL;
XX      Santos Garcia MA, Seubberger H;
PI

```

```

XX  Bennett CF, Cowser LM, Condon TP;
PI  WPI: 1999-590416/50.
XX  P-PSDB: AA44154.
DR  Antisense inhibition of integrin alpha4 expression useful for treating
XX  inflammatory diseases such as atherosclerosis, allergies, asthma and
XX  arthritis -
XX  Example 12; Column 71-74; 40pp; English.
XX  This sequence represents the coding region of the mouse integrin
XX  alpha4 gene. The invention relates to the generation of antisense
XX  oligonucleotides targeted to the integrin alpha4 gene which are used
XX  for inhibiting expression of the integrin alpha4 mRNA or protein.
XX  Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
XX  called alpha4beta1 and CD49g/CD29). VLA-4 is expressed on the cell
XX  surfaces of leucocytes and vascular endothelial cells and mediates the
XX  adhesion of leucocytes to the vascular endothelium prior to migration
XX  into the surrounding tissues. This migration is an essential step in
XX  inflammation and hence VLA-4 (and consequently integrin alpha4) is a
XX  potential therapeutic target for treating inflammatory diseases and
XX  the damaging effects of excessive inflammation. These disorders include
XX  atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
XX  metastasis (VLA-4 is involved in migration of the tumor cells through
XX  the extracellular matrix into the circulatory system). VLA-4 is also
XX  involved in a number of autoimmune diseases such as Grave's disease,
XX  Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis.
XX  VLA-4 may also be involved in promoting adhesion (i.e. retention) of
XX  hemopoietic stem cells in bone-marrow and in allograft rejection.
XX  Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other:
SQ

Query Match          64.8%; Score 16.2; DB 20; Length 1771;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY  4 attgtcgcaagtcacagaatt 24
    1 11 11111111111111111111
DB  306 agtggagcgaagtcacagaatt 326

RESULT 29
AAS10264
ID  AAS10264 standard; DNA; 1771 BP.
XX
XX  AAS10264:
AC  24-OCF-2001 (first entry)
DT  Mouse integrin alpha 4 partial gene sequence.
XX
XX  Mouse; integrin alpha 4; antisense; L20788; very late antigen 4; VLA4;
XX  autoimmune disease; inflammatory disease; rheumatoid arthritis;
XX  multiple sclerosis; tumor metastasis; melanoma; asthma; psoriasis;
XX  allergy; Grave's disease; Hashimoto's thyroiditis;
XX  systemic lupus erythematosus; allograft rejection; ds.
XX
XX  Mus musculus.
OS
XX
XX  Key Location/Qualifiers
XX  CDS 1193..1771
XX  FT /*tag= a
XX  FT /product= "Integrin alpha 4"
XX  FT /partial
XX  FT /note= "No stop codon"
XX  FT intron
XX  FT /*tag= b
XX  FT /number= 1
XX
XX  US6258790-B1.

```

```

XX  10-JUL-2001.
XX  19-AUG-1999; 99US-0377309.
XX  05-OCF-1998; 98US-0166203.
XX  (ISIS-) ISIS PHARM INC.
XX  Bennett CF, Condon TP, Cowser LM;
PI  WPI: 2001-450381/48.
XX  P-PSDB: AA04967.
XX  Composition for treating inflammatory and autoimmune diseases,
XX  comprising antisense compound targeted to nucleic acid molecule encoding
XX  integrin alpha4 and inhibit expression of integrin alpha4 -
XX  Example 12; Column 77-79; 49pp; English.
XX  The sequence (Genbank accession number L20788) is the partial gene
XX  sequence of mouse integrin alpha 4, one component of very late antigen 4
XX  (VLA4) which is involved in various autoimmune and inflammatory diseases.
XX  The invention relates to antisense inhibitors of integrin alpha 4 which
XX  target and inhibit expression of integrin alpha 4. The antisense
XX  molecules are useful for inhibiting the expression of integrin alpha4 in
XX  human cells or tissues, treating an animal having a disease or
XX  condition associated with expression of integrin alpha4, e.g.,
XX  inflammatory disease or condition, autoimmune disease or condition
XX  including rheumatoid arthritis, multiple sclerosis and tumor
XX  metastases, melanoma, asthma, psoriasis, allergy, Grave's disease,
XX  Hashimoto's thyroiditis, systemic lupus erythematosus and allograft
XX  rejection, and diseases or conditions characterised by leukocyte
XX  migration into affected tissues, preferably central nervous system
XX  tissues. The antisense molecules are also useful for reducing the
XX  levels of VLA-4 and alpha4beta1 integrin in human cells or tissues,
XX  and reducing the adherence of cells of a first type e.g., melanoma cells
XX  or lymphocytes, to cells of a second type e.g., endothelial cells, by
XX  inhibiting integrin alpha4 expression and thus decreasing adhesion of
XX  cells.
XX  Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other:
SQ

Query Match          64.8%; Score 16.2; DB 22; Length 1771;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY  4 attgtcgcaagtcacagaatt 24
    1 11 11111111111111111111
DB  306 agtggagcgaagtcacagaatt 326

RESULT 30
AA051236
ID  AA051236 standard; CDNA; 2114 BP.
XX
XX  AA051236:
AC  11-MAY-1994 (first entry)
DT  Plant NADPH cytochrome P450 reductase (ara B).
XX
XX  Plant NADPH cytochrome P450 reductase (ara B).
XX  NADPH cytochrome P450 reductase; functional complementation;
XX  identification; ss.
XX  Arabidopsis thaliana.
OS
XX
XX  Key Location/Qualifiers
XX  CDS 36..2114
XX  FT /*tag= a
XX  FT /product= NADPH cytochrome P450 reductase.
XX  FT misc_difference 510..512
XX
XX  US6258790-B1.

```

```

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

```

```

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157863.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      64.8%; Score 16.2; DB 21; Length 1150;
Best local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 gacatgtctgcagatcacaga 21
        ||||| ||||| ||| |||
Db       79 GACATGATGCGCAACTCAAGA 59

```

```

RESULT 28
AA30602
ID AA230602 standard; DNA; 1771 BP.
XX
AC AA230602;
XX
DT 18-JAN-2000 (first entry)
XX
DE Mouse integrin alpha 4 coding sequence.
XX
KW Human; integrin; antisenase; oligonucleotide; inhibition; expression;
very late antigen; CD29; cell surface; leucocyte; adhesion;
vascular endothelial cell; vascular endothelium; migration; inflammation;
atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
metastasis; circulatory system; autoimmune disease; Grave's disease;
Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS Mus sp.
XX
PN US5968826-A.
XX
PD 19-OCT-1999.
XX
PF 05-OCT-1998; 98US-0166203.
XX
PR 05-OCT-1998; 98US-0166203.
XX
PA (ISIS-) ISIS PHARM INC.

```


tissue. They can be used as modulating agents for regulating cellular processes, thus, the proteins and their associated nucleic acids can be used to prognosticate, prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention deficit disorder, Crohn's disease, gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and muscular dystrophy. Antibodies to disorders such as these can be made by providing a polypeptide of the invention to an immuno-competent vertebrate and harvesting blood or serum from the vertebrate.

Sequence 4074 BP: 1280 A; 811 C; 854 G; 1120 T; 9 other;

Query Match 66.4%; Score 16.6; DB 22; Length 4074;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacatgtcgcaagtcacagaatt 24
||||| ||||| ||||| |||||
DB 3509 gaattmtagcaagccactgaatt 3532

RESULT 24
AA068433/C
ID AA068433 standard; DNA: 5198 BP.

AC AA068433;
DT 26-JAN-1995 (first entry)
XX
XX

DE Maize Hm1 gene.

KW Hm1 gene; HC toxin; race-specific resistance; disease-resistance;
KM Cochliobolus carbonum; selectable marker; transformation;
transposon tagging; ds.

XX
XX

OS Zea mays.

PN W09413825-A.

PD 23-JUN-1994.

PF 14-DEC-1993; 93MO-US12146.

PR 15-DEC-1992; 92US-0995658.

XX
XX

PA (PION-) PIONEER HI-BRED INT INC.

XX
XX

PI Briggs SP, Johal GS;

DR WPI: 1994-217898/26.

XX
XX

PT Hm1 gene conferring race-resistance to Cochliobolus carbonum to

PT cells - for use as a selectable marker for transformed maize

XX
XX

PS Disclosure: Page 14-16; 19pp; English.

XX
XX

CC Transposon mutagenesis was used to tag, clone and characterize the

CC maize Hm1 gene. Genomic and cDNA sequences of the Hm1 gene are

CC provided in AA068433 and AA068434, respectively.

XX
XX

DB 4891 AAATTATCGCAAGTCATTAATAATT 4869
||||| ||||| ||||| |||||
RESULT 25
AA099462/C
ID AA099462 standard; DNA: 5198 BP.
AC AA099462;
DT 30-DEC-1995 (first entry)
XX
XX
DE Maize Hm1 genomic DNA.
KM Hm1 gene; Cochliobolus carbonum Nelson race 1; fungus;
KM fungal disease toxin; resistance; ss.
XX
XX
OS Zea mays.
PN W09507989-A.
XX
PD 23-MAR-1995.
XX
PF 19-SEP-1994; 94MO-US10497.
XX
PR 17-SEP-1993; 93US-0123761.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX
PI Briggs SP, Johal GS;
DR WPI: 1995-131357/17.
XX
PT Disease resistance gene Hm1 from maize - is used to confer
PT resistance to Cochliobolus carbonum.
XX
PS Claim 2; Page 20-22; 26pp; English.
XX
CC A DNA sequence is claimed having 90% translational homology to
CC genomic or cDNA sequence of maize Hm1 gene (AA099462 & AA099463). The
CC gene Hm1 controls resistance to Cochliobolus carbonum Nelson race 1.
CC The Hm1 gene can be used in conjunction with HC-toxin in a
CC selectable marker system for use in maize transformation.
CC This gene can also be inserted into the genomes of maize varieties
CC which lack the gene to produce transformants that are resistance
CC to disease caused by C. carbonum. The Hm1 resistance gene is one of
CC a family of homologous disease resistance genes in maize and other
CC crops. The invention encompasses methods for identification of
CC homologous disease resistance genes by using this gene or a
CC fragment of it as a probe.
XX
XX
SO Sequence 5198 BP: 1375 A; 1159 C; 1136 G; 1528 T; 0 other;

Query Match 66.4%; Score 16.6; DB 16; Length 5198;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acatgtcgcaagtcacagaatt 24
||||| ||||| ||||| |||||
DB 4891 AAATTATCGCAAGTCATTAATAATT 4869

RESULT 26
AAC24233/C
ID AAC24233 standard; cDNA: 386 BP.
XX
XX

AC AAC24233;

DT 06-OCT-2000 (first entry)

XX
XX

DE Human secreted protein 5' EST, SEQ ID NO: 28308.

CC immuno-competent vertebrate and harvesting blood or serum from the
CC vertebrate.

SQ Sequence 3985 BP; 1258 A; 788 C; 829 G; 1101 T; 9 other;

Query Match 66.4%; Score 16.6; DB 22; Length 3985;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gacattgctgcaagtcacagaatt 24
Db 3423 gaattttagcaagcactagaatt 3446

RESULT 22

AAS00662 ID AAS00662 standard; cDNA: 4018 BP.

AC AAS00662;

DT 07-SEP-2001 (first entry)

DE Human INTERCEPT 289 form 1b cDNA.

XX Human; INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anapylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.

XX Homo sapiens.

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 179..715

FT /tag= a

FT /product= "Human INTERCEPT 289 form 1b"

PN WO200129088-A1.

PD 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI; 2001-308477/32.

XX P-PSDB; AAU00672.

XX New isolated nucleic acid molecule for diagnosis, prevention, and

XX therapy of human and other animal disorder, or as modulating agent for

XX regulating cellular processes -

XX Claim 1; Fig 2D-2G; 263pp; English.

XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 1b
CC polypeptide. This protein and similar others exhibit the ability to
CC affect growth, proliferation, survival, differentiation, activity,
CC morphology, or movement/migration of, e.g. T cells and cells of the
CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
CC tissue. They can be used as modulating agents for regulating cellular
CC processes, thus, the proteins and their associated nucleic acids can be
CC used to prophylactically prevent, diagnose, or treat disorders associated
CC with physiological processes. These disorders include abnormal blood
CC coagulation, asthma, anapylaxis, hepatitis, multiple sclerosis, cancer,

CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
CC as these can be made by providing a polypeptide of the invention to an
CC immuno-competent vertebrate and harvesting blood or serum from the
CC vertebrate.

SQ Sequence 4018 BP; 1250 A; 807 C; 844 G; 1108 T; 9 other;

Query Match 66.4%; Score 16.6; DB 22; Length 4018;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gacattgctgcaagtcacagaatt 24
Db 3456 gaattttagcaagcactagaatt 3479

RESULT 23

AAS00661 ID AAS00661 standard; cDNA: 4074 BP.

AC AAS00661;

DT 07-SEP-2001 (first entry)

DE Human INTERCEPT 289 form 1a cDNA.

XX Human; INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anapylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.

XX Homo sapiens.

Key Location/Qualifiers

FT CDS 179..745

FT /tag= a

FT /product= "Human INTERCEPT 289 form 1a"

PN WO200129088-A1.

PD 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI; 2001-308477/32.

XX P-PSDB; AAU00671.

XX New isolated nucleic acid molecule for diagnosis, prevention, and

XX therapy of human and other animal disorder, or as modulating agent for

XX regulating cellular processes -

XX Claim 1; Fig 2A-2C; 263pp; English.

XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 1a
CC polypeptide. This protein and similar others exhibit the ability to
CC affect growth, proliferation, survival, differentiation, activity,
CC morphology, or movement/migration of, e.g. T cells and cells of the
CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus

Qy 1 gacattgtcgaagtcacagaatt 24
 || ||| | ||||| ||| |||||
 Db 3363 gaattntagcaagcactgaatt 3386

RESULT 20

AAS00664
 ID AAS00664 standard; cDNA: 3958 BP.
 XX
 AC AAS00664;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Human INTERCEPT 289 form 2b cDNA.

Human; INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
 skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
 malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.

OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT 162..629
 FT CDS /*tag= a
 FT /product= "Human INTERCEPT 289 form 2b"

W0200129088-A1.

26-APR-2001.

23-JUN-2000; 2000MO-US17386.

19-OCT-1999; 99US-0420707.

(MILL-) MILLENNIUM PHARM INC.

Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

WPI: 2001-308477/32.

P-PSDB; AAU00674.

New isolated nucleic acid molecule for diagnosis, prevention, and
 therapy of human and other animal disorder, or as modulating agent for
 regulating cellular processes -

Claim 1: Fig 2b-20; 263bp; English.

The sequence represents a cDNA which encodes human INTERCEPT 289 form 2b
 polypeptide. This protein and similar others exhibit the ability to
 affect growth, proliferation, survival, differentiation, activity,
 morphology, or movement/migration of, e.g. T cells and cells of the
 heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
 spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
 tissue. They can be used as modulating agents for regulating cellular
 processes, thus, the proteins and their associated nucleic acids can be
 used to prognosticate, prevent, diagnose, or treat disorders associated
 with physiological processes. These disorders include abnormal blood
 coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
 coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
 sclerosis, meningitis, attention deficit disorder, Crohn's disease,
 gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
 pulmonary embolism and muscular dystrophy. Antibodies to disorders such
 as these can be made by providing a polypeptide of the invention to an
 immuno-competent vertebrate and harvesting blood or serum from the
 vertebrate.

Sequence 3958 BP; 1234 A; 796 C; 829 G; 1090 T; 9 other:

Query Match 66.4%; Score 16.6; DB 22; Length 3958;
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtcacagaatt 24
 || ||| | ||||| ||| |||||
 Db 3370 gaattntagcaagcactgaatt 3393

RESULT 21

AAS00663
 ID AAS00663 standard; cDNA: 3985 BP.
 XX
 AC AAS00663;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Human INTERCEPT 289 form 2a cDNA.

Human; INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
 skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
 malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.

OS Homo sapiens.

XX
 FH Key location/Qualifiers
 FT 162..659
 FT CDS /*tag= a
 FT /product= "Human INTERCEPT 289 form 2a"

W0200129088-A1.

26-APR-2001.

23-JUN-2000; 2000MO-US17386.

19-OCT-1999; 99US-0420707.

(MILL-) MILLENNIUM PHARM INC.

Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

WPI: 2001-308477/32.

P-PSDB; AAU00673.

New isolated nucleic acid molecule for diagnosis, prevention, and
 therapy of human and other animal disorder, or as modulating agent for
 regulating cellular processes -

Claim 1: Fig 2b-2K; 263bp; English.

The sequence represents a cDNA which encodes human INTERCEPT 289 form 2a
 polypeptide. This protein and similar others exhibit the ability to
 affect growth, proliferation, survival, differentiation, activity,
 morphology, or movement/migration of, e.g. T cells and cells of the
 heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
 spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
 tissue. They can be used as modulating agents for regulating cellular
 processes, thus, the proteins and their associated nucleic acids can be
 used to prognosticate, prevent, diagnose, or treat disorders associated
 with physiological processes. These disorders include abnormal blood
 coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
 coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
 sclerosis, meningitis, attention deficit disorder, Crohn's disease,
 gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
 pulmonary embolism and muscular dystrophy. Antibodies to disorders such
 as these can be made by providing a polypeptide of the invention to an

```

XX 07-SEP-2001 (first entry)
XX
XX Human INTERCEPT 289 form 3b cDNA.
DE
XX Human: INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 162..569
FT /tag= a
FT /product= "Human INTERCEPT 289 form 3b"
XX
XX WO200129088-A1.
PN
XX
XX 26-APR-2001.
XX
XX 23-JUN-2000; 2000WO-US17386.
XX
XX 19-OCT-1999; 99US-0420707.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
PI
XX WPI: 2001-308477/32.
DR
XX P-PSDB; AAU00676.
XX
XX New isolated nucleic acid molecule for diagnosis, prevention, and
PT therapy of human and other animal disorder, or as modulating agent for
PT regulating cellular processes -
XX
XX Claim 1; Fig 2P-2V; 263pp; English.
XX
XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 3b
CC polypeptide. This protein and similar others exhibit the ability to
CC affect growth, proliferation, survival, differentiation, activity,
CC morphology, or movement/migration of, e.g., T cells and cells of the
CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
CC tissue. They can be used as modulating agents for regulating cellular
CC processes, thus, the proteins and their associated nucleic acids can be
CC used to prognosticate, prevent, diagnose, or treat disorders associated
CC with physiological processes. These disorders include abnormal blood
CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
CC as these can be made by providing a polypeptide of the invention to an
CC immuno-competent vertebrate and harvesting blood or serum from the
CC vertebrate.
XX
XX Sequence 3898 BP; 1213 A; 781 C; 818 G; 1077 T; 9 other;
SO

```

Query Match 66.4%; Score 16.6; DB 22; Length 3898;
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

RESULT 19
XX AAS00665
XX ID AAS00665 standard; cDNA; 3925 BP.
XX
XX AAS00665;
AC
XX
XX 07-SEP-2001 (first entry)
XX
XX Human INTERCEPT 289 form 3a cDNA.
DE
XX Human: INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 162..599
FT /tag= a
FT /product= "Human INTERCEPT 289 form 3a"
XX
XX WO200129088-A1.
PN
XX
XX 26-APR-2001.
XX
XX 23-JUN-2000; 2000WO-US17386.
XX
XX 19-OCT-1999; 99US-0420707.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
PI
XX WPI: 2001-308477/32.
DR
XX P-PSDB; AAU00675.
XX
XX New isolated nucleic acid molecule for diagnosis, prevention, and
PT therapy of human and other animal disorder, or as modulating agent for
PT regulating cellular processes -
XX
XX Claim 1; Fig 2P-2S; 263pp; English.
XX
XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 3a
CC polypeptide. This protein and similar others exhibit the ability to
CC affect growth, proliferation, survival, differentiation, activity,
CC morphology, or movement/migration of, e.g., T cells and cells of the
CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
CC tissue. They can be used as modulating agents for regulating cellular
CC processes, thus, the proteins and their associated nucleic acids can be
CC used to prognosticate, prevent, diagnose, or treat disorders associated
CC with physiological processes. These disorders include abnormal blood
CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
CC as these can be made by providing a polypeptide of the invention to an
CC immuno-competent vertebrate and harvesting blood or serum from the
CC vertebrate.
XX
XX Sequence 3925 BP; 1237 A; 773 C; 818 G; 1088 T; 9 other;
SO

```

Query Match 66.4%; Score 16.6; DB 22; Length 3925;
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142970.
PR 13-JUL-1999; 99US-0142972.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144842.
PR 21-JUL-1999; 99US-0144844.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.4%; Score 16.6; DB 21; Length 3717;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacatgtcgaagtcacagaat 23
||||| ||||| |||||
Db 2935 gacatgttcgaagtcagaat 2957

RESULT 18
AAS00666
ID AAS00666 standard; CDNA; 3898 BP.
XX
AC AAS00666;

Query Match 66.4%; Score 16.6; DB 8; Length 2157;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 catgttcgcaagtcacagaatta 25
||||| ||| ||| |||
DB 958 CATGTGCTAGTACGACAGACTAA 936

RESULT 16

AAH20884
ID AAH20884 standard; DNA: 2421 BP.

AC AAH20884;

DT 24-AUG-2001 (first entry)

DE H. polymorpha 25S rDNA encoding DNA.

KM 25S rDNA; integration vector; recombinant protein production;
KW DNA integration; ds.

OS Hansenula polymorpha.

PN WO200138510-A2.

PD 31-MAY-2001.

PF 23-NOV-2000; 2000WO-EP11687.

PR 23-NOV-1999; 99DE-1056297.

PR 03-DEC-1999; 99DE-1058327.

PR 29-DEC-1999; 99DE-1063690.

PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.

PI Gellissen G, Diesel A, Klabunde J, Suckow M, Hollenberg CP;

DR WPI; 2001-355926/37.

PT New integration vector for protein production, useful e.g. for
PT expressing insulin, contains a yeast rDNA sequence, selection marker
PT and expression cassette

PS Claim 15a; Fig 2; 54pp; German.

CC This invention describes a novel integration vector (A) for expressing
CC at least one protein (I) in a host cell (especially a fungus) comprises
CC at least one each of: (1) rDNA sequence (II) from yeast; (2) a
CC non-deficient selection marker gene (III); and (3) an expression cassette
CC (EC) which comprises a promoter and terminator functional in the host
CC organism and the (1)-encoding region of a heterologous or endogenous
CC gene. (A) are used for production of recombinant proteins, especially
CC in filamentous fungi or yeast e.g. phytase, insulin and reporter
CC proteins. (A) provide targeted and stable integration of heterologous
CC DNA (including several different genes for simultaneous co-expression)
CC into a host cell genome, optionally at high copy number. No deficient
CC selection markers are required and several vectors, containing the same
CC non-deficient marker, can be integrated simultaneously, with retention
CC of mitotic stability. This sequence represents the Hansenula polymorpha
CC 25S rDNA encoding sequence described in the method of the invention.

XX Sequence 2421 BP; 712 A; 622 C; 343 G; 741 T; 3 other;

Query Match 66.4%; Score 16.6; DB 22; Length 2421;
Best Local Similarity 79.2%; Pred. No. 1.4e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgtcgaagtcacagaatt 24
||||| ||| ||| |||
DB 616 gacattgtcgaagtcacagaatt 639

RESULT 17

AAH48934
ID AAC48934 standard; DNA: 3717 BP.

AC AAC48934;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment seq ID NO: 59311.

KM Hybridisation assay; genetic mapping; gene expression control;
KM Protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

KM	Hml gene; HC toxin; race-specific resistance; disease-resistance
KM	Cochliobolus carbonum; selectable marker; transformation;
KM	transposon tagging; ds.
XX	
OS	Zea mays.
XX	
PN	MO9413825-A.
XX	
PD	23-JUN-1994.
XX	
PF	14-DEC-1993; 93MO-US12146.
XX	
PR	15-DEC-1992; 92US-0995658.
XX	
PA	(PION-) PIONEER HI-BRED INT INC.
XX	
PI	Briggs SP, Johal GS;
XX	
DR	WPI; 1994-217898/26.
XX	
PT	Hml gene conferring race-resistance to Cochliobolus carbonum to
PT	maize - for use as a selectable marker for transformed maize
XX	cells
XX	
PS	Disclosure: Page 18; 19pp; English.
XX	
CC	Transposon mutagenesis was used to tag, clone and characterize the
CC	maize Hml gene. Genomic and cDNA sequences of the Hml gene are
CC	provided in AA068433 and AA068434, respectively.
XX	
SO	Sequence 1374 BP; 285 A; 406 C; 432 G; 251 T; 0 other;

Query Match	66.4%	Score 16.6; DB 15;	Length 1374;
Best Local Similarity	82.6%	Pred. No. 1.3e+02;	
Matches 19; Conservative	0;	Mismatches 4;	Indels 0; Gaps 0;
QY	2 acattgtcgcagtcacagaatt 24		
Db	1329 AAATTATCGCACTCTATAAATT 1307		

RESULT	14
AAQ99463/c	
ID	AAQ99463 standard; cDNA; 1374 BP.
XX	
AC	AAQ99463;
XX	
DT	30-DEC-1995 (first entry)
XX	
DE	Maize Hm1 gene cDNA.
XX	
KW	Hm1 gene; Cochliobolus carbonum Nelson race 1; fungus;
XX	
OS	fungal disease toxin; resistance; ss.
XX	
PN	Zea mays.
XX	
PN	W09507989-A.
XX	
PD	23-MAR-1995.
XX	
PF	19-SEP-1994; 94MO-US10497.
XX	
PR	17-SEP-1993; 93US-0123761.
XX	
PA	(PION-) PIONEER HI-BRED INT INC.
XX	
PI	Briggs SP, Johal GS;
XX	
DR	WPI; 1995-131357/17.
PT	
PT	Disease resistance gene Hm1 from maize - is used to confer resistance to Cochliobolus carbonum.

XX Claim 2: Page 23: 26pp: English.

PS

XX

CC A DNA sequence is claimed having 90% translational homology to

CC genomic or cDNA sequence of maize Hm1 gene (AA099462 & AA099463). The

CC gene Hm1 controls resistance to *Cochliobolus carbonum* Nelson race 1.

CC The Hm1 gene can be used in conjunction with HC-toxin in a

CC selectable marker system for use in maize transformation.

CC This gene can also be inserted into the genomes of maize varieties

CC which lack the gene to produce transformants that are resistance

CC to disease caused by *C. carbonum*. The Hm1 resistance gene is one of

CC a family of homologous disease resistance genes in maize and other

CC crops. The invention encompasses methods for identification of

CC homologous disease resistance genes by using this gene or a

CC fragment of it as a probe.

XX

XX

50 Sequence 1374 BP; 287 A; 405 C; 432 G; 250 T; 0 other:

Query Match:	66.4%	Score 16.6:	DB 16:	Length 1374:
Best Local Similarity:	82.6%	Pred. No. 1.3e+02:		
Matches 19:	Conservative	0:	Mismatches 4:	Indels 0:
Gaps	0:			
2	acattgtcgcagatcacagaatt	24		
1				
1329	AAATTATCGCAGACGATTAATAATT	1307		

RESULT	15
ID	AAAT70438/c
XX	
AC	AAAT70438;

FH	Key	Location/Qualifiers
FT		40.1219
FT	CDS	/*tag= a
FT		/label=SACP gene
FT		

CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.

XX Sequence 322 BP; 115 A; 42 C; 41 G; 115 T; 9 other;

Query Match 68.8%; Score 17.2; DB 16; Length 322;
Best Local Similarity 86.4%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 catgtcgcaagtcacagaatt 24
||||| 11 11 1111111111
DB 5 catctcgcaattcacaagaatt 26

RESULT 11
AA256756/c
ID AA256756 standard; cDNA: 902 BP.

XX AA256756;

DT 23-MAR-2000 (first entry)

XX Human transmembrane protein HTMPN-59 encoding cDNA.

XX Human; transmembrane protein; HTMPN; diagnosis; immunospecific;

KW antiproliferative; neuroprotective; immune disorder;

KW reproductive disorder; smooth muscle disorder; neurological disorder;

KW gastrointestinal disorder; developmental disorder;

KW cell proliferative disorder; ss.

XX Homo sapiens.

XX WO9961471-A2.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US11904.

XX 29-MAY-1998; 98US-0087260.

PR 02-JUL-1998; 98US-0091674.

PR 02-OCT-1998; 98US-0102954.

PR 24-NOV-1998; 98US-0109869.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;

PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;

PI Au-Young J;

XX WPI: 2000-072605/06.

DR P-PSDB; AAY57935.

XX Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders -

XX Claim 9; Page 215; 229pp; English.

XX AA256698 to AA256776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.

CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,

CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or

CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The

CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.

CC Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 902;
Best Local Similarity 86.4%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 atgtcgcaagtcacagaatta 25
||||| 11 11 1111111111
DB 434 AATGGCGCTAGTCACAGCAATTA 413

RESULT 12
AA056307/c
ID AA056307 standard; DNA: 17 BP.

XX AA056307;

DT 13-APR-1994 (first entry)

XX BioB DNA primer.

XX BioA; BioB; promoter; biotin; operon; primer; ss.

XX Synthetic.

XX JP05219956-A.

XX 31-AUG-1993.

XX 14-SEP-1992; 92UP-0244792.

XX 14-SEP-1992; 92UP-0244792.

XX (SHIS) SHISEIDO CO LTD.

PA (TAKE) TAKEDA CHEM IND LTD.

XX WPI: 1993-308323/39.

XX DNA sequence of biotin operon - has base sequence of E. coli

PT mutated by base pair(s) compared to wild type

XX Example 1; Fig 8; 11pp; Japanese.

CC A novel DNA sequence comprises the E. coli biotin operon (BO) in which
CC the control region of BO or the region near the bioB initiation
CC codon is mutated by at least one base pair compared to its

CC wild type. Two primers (AA056306-Q56307) are described in Example 1.

CC A microorganism belonging to Escherichia genus, transformed by

CC a recombinant plasmid carrying such DNA can be used for the prodn.

CC of biotin-active substances.

XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gtgcgaagtcacagaat 23
||||| 11 11 1111111111
DB 17 GTGCAAGTCACAGAAAT 1

RESULT 13
AA068434/c
ID AA068434 standard; cDNA: 1374 BP.

XX AA068434;

DT 26-JAN-1995 (first entry)

XX Maize hml cDNA.

AC	AAO62386:
XX	16-NOV-1994 (first entry)
DT	Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX	
XX	Biotin: expression; enterobacteria; vitamin H; synthesis;
KW	plasmid: pB030A-15/9; bioB: bioF: bioC: bioD: bioA;
KW	promoter plac: biotin synthase; KAPA synthase;
KW	8-amino-7-oxononanoate synthase; pimeloyl-CoA: DTB synthase;
KW	deshiobiotin synthase; DAPA synthase;
KX	s-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX	seborrhoea: dermatitis; ds.
OS	Escherichia coli DSM498.
XX	
FH	Location/Qualifiers
FT	Key
FT	promoter
FT	1..96
FT	/tag= a
FT	/function= "promoter plac"
FT	/evidence= EXPERIMENTAL
FT	23..28
FT	-35_signal
FT	/tag= b
FT	/standard_name= "promoter plac"
FT	45..50
FT	-10_signal
FT	/tag= c
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "promoter plac"
FT	105..109
FT	RBS
FT	/tag= d
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "bioB RBS no. 9"
FT	CDS
FT	117..1157
FT	/tag= e
FT	/product= "biotin synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioB"
FT	/number= 1
FT	RBS
FT	1141..1146
FT	/tag= f
FT	/standard_name= "bioF RBS"
FT	CDS
FT	1154..2311
FT	/tag= g
FT	/EC_number= 2.3.1.47
FT	/product= "KAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioF"
FT	/number= 2
FT	/standard_name= "8-amino-7-oxononanoate synthase"
FT	RBS
FT	2284..2288
FT	/tag= h
FT	/standard_name= "bioC RBS"
FT	CDS
FT	2295..3050
FT	/tag= i
FT	/function= "involved in pimeloyl-CoA synthesis"
FT	/product= "protein"
FT	/gene= "bioC"
FT	/number= 3
FT	RBS
FT	3030..3033
FT	/tag= j
FT	/standard_name= "bioD RBS"
FT	CDS
FT	3043..3753
FT	/tag= k
FT	/EC_number= 6.3.3.3
FT	/product= "DTB synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioD15"
FT	/number= 4
FT	/standard_name= "deshiobiotin synthase"
FT	3712..3750
FT	/tag= l
FT	/note= "bioD15 substitution"
FT	3742..3746
FT	misc_RNA
FT	RBS

FT		/tag= m
FT	/standard_name= "DioA RBS"	
FT	3750..5039	
CDS		
FT	/tag= n	
FT	/EC_number= 2.6.1.62	
FT	/product= "DAPA synthase"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "dioA"	
FT	/number= 5	
FT	/standard_name= "S-adenosyl-L-methionine: 8-amino-	
FT	-7-oxononanotate aminotransferase"	
RBS		
FT	5088..5093	
FT	/tag= O	
FT	/standard_name= "ORF1 RBS"	
FT	5098..5574	
CDS		
FT	/tag= P	
FT	/function= "unknown, involved in biotin synthesis"	
FT	/product= "protein"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "ORF1"	
FT	/number= 6	
terminator		
FT	5583..5644	
FT	/tag= q	
FT	/standard_name= "rho-independent transcriptional	
FT	terminator"	
stem_loop		
FT	5583..5605	
FT	/tag= r	
PX		
PX	WQ9408023-A.	
PX		
PX	14-APR-1994.	
PX		
PE	01-OCT-1993:	93WO-EPO2688.
PX		
PR	02-OCT-1992:	92CH-0003124.
PR	15-JUL-1993:	93CH-0002134.
PX		
(LONZ) LONZA AG.		
PA		
PX		
PL	Birch O., Brass J., Fuhrmann M., Shaw N;	
PX		
DR	WPI: 1994-135587/16.	
DR	P-PsDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR51887.	
PX		
PT	Biotechnological biotin prodn. using enterobacterial biotin-gene	
PT	- providing vitamin H in high yield	
PX		
PS	Claim 1: Fig 6, Page 47-55 and 60-65; 92pp; German.	
PX		
CC	The sequence is derived from plasmid pB030A-15/9 contg. the	
CC	bioB, bioF, bioC, bioD and bioG genes responsible for biosynthesis	
CC	of biotin, arranged in a transcription unit. Microorganisms	
CC	contg. these DNA fragments or plasmids may be used in the prodn.	
CC	of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,	
PX	loss of appetite and tiredness.	
SQ		
SQ	Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;	
Query Match	100.0%; Score 25; DB 15; Length 5872;	
Best Local Similarity	100.0%; Pred. No. 0.016;	
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY		
OY	1 gacattgtcgaagtcacagaatta 25	
OY		
OY		
Dd	137 gacatttctcgcaagtcatcacagaatta 161	
RESULT	6	
AHH68531		
ID	AHH68531 standard; DNA; 349980 BP.	
PX		
PX	AHH68531;	

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA;
 XX
 DR WPI; 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di- amino- pelarmonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2; Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BiotB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatgtcgaagtcacagaatta 25
 ||||||||||||||||||||
 Db 21 gacatgtcgaagtcacagaatta 45

RESULT 3
 ID AAN91329 standard; DNA; 1084 BP.
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KM E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI; 1989-295085/41. P-PSDB P91392
 XX
 PT plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatgtcgaagtcacagaatta 25
 ||||||||||||||||||||
 Db 44 gacatgtcgaagtcacagaatta 68

RESULT 4
 ID AAN60496 standard; DNA; 1121 BP.
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KM Biotin synthetic enzyme; E.coli; deschiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag= a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI; 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure: Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing deschiobiotin.
 CC
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatgtcgaagtcacagaatta 25
 ||||||||||||||||||||
 Db 62 gacatgtcgaagtcacagaatta 86

RESULT 5
 ID AAO62386 standard; DNA; 5872 BP.
 XX

85 15.6 62.4 721 22 AAH41919 Human CLASP-3 geno
86 15.6 62.4 1002 19 AAV30946 Staphylococcus aur
87 15.6 62.4 1020 19 AAV30947 Staphylococcus aur
88 15.6 62.4 1077 21 AAC75310 Canine angiotensin
C 89 15.6 62.4 1276 21 AAC74265 Human secreted pro
C 90 15.6 62.4 1542 22 AAI62949 Human genomic DNA
C 91 15.6 62.4 1585 20 AAZ41281 Human normal ovari
C 92 15.6 62.4 1676 22 AAH34149 Human colon cancer
C 93 15.6 62.4 1790 22 AAH98422 Human EST-derived
C 94 15.6 62.4 1905 21 AAC79713 Human secreted pro
C 95 15.6 62.4 1953 20 AAC20372 Nucleotide sequenc
C 96 15.6 62.4 1953 22 AAF89034 C elegans FATPa co
C 97 15.6 62.4 2001 22 AAH33500 Human colon cancer
C 98 15.6 62.4 2533 22 AAH18302 Human CDNA sequenc
C 99 15.6 62.4 2730 21 AAC76029 Human ORFX ORF1584
100 15.6 62.4 3360 19 AAV30948 Staphylococcus aur

ALIGNMENTS

RESULT 1

AAI62941 ID AAI62941 standard; DNA: 839 BP.

AC AAI62941:

XX 22-OCT-2001 (first entry)

DT Human genomic DNA SEQ ID NO 269.

XX Human: nontropic; neuroprotective; cytoskeletal; dermatological; virocidic;
KM immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
KM antiparasitism; antitickling; antianemic; antirheumatic; cancer;
KM antihemetic; hepatotropic; cerebroprotective; antineoplastic; antitumor;
KM anticancer; anticancer; anticancer; anticancer; anticancer; anticancer;
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KM neurological disease; infection; nephrotropic; gene therapy; vaccine;
ds.

XX Homo sapiens.

XX WO200155449-A1.

XX 02-AUG-2001.

PF 17-JAN-2001: 2001MO-US01346.

XX 31-JAN-2000: 2000US-0179065.
PR 04-FEB-2000: 2000US-0180628.
PR 19-MAY-2000: 2000US-020515.
PR 07-JUL-2000: 2000US-0216880.
PR 14-JUL-2000: 2000US-0218290.
PR 14-AUG-2000: 2000US-0225447.
PR 01-SEP-2000: 2000US-0229343.
PR 06-SEP-2000: 2000US-0230437.
PR 08-SEP-2000: 2000US-0231243.
PR 25-SEP-2000: 2000US-0234997.
PR 29-SEP-2000: 2000US-0236367.
PR 13-OCT-2000: 2000US-0239337.
PR 08-NOV-2000: 2000US-0246476.
PR 08-NOV-2000: 2000US-0246477.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246528.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249265.
PR 01-DEC-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.

PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM:

DR WPI: 2001-476225/51.

XX Novel plasma membrane associated proteins useful for diagnosing,

PT treating, preventing and/or prognosing disorders related to the

PT proteins, including cancer, immune response and neuronal disorders -

XX Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.

XX The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAI62347-AAI62415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, Rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPo at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 839 BP: 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;

Best Local Similarity 100.0%; Pred. No. 0.013;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaatta 25

DB 53 gacattgcgaagtcacagaatta 77

RESULT 2

AAAX01303 ID AAAX01303 standard; DNA: 1041 BP.

XX AAAX01303:

DT 12-APR-1999 (first entry)

XX E. coli biotin synthetase (BioB) coding sequence.

XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;

KW biotin synthase; biotin production; vitamin H; BioB; ss.

XX Escherichia coli.

OS US5869719-A.

PN 09-FEB-1999.

XX 30-APR-1997; 97US-0846338.

XX 30-APR-1997; 97US-0846338.

XX 08-MAR-1995; 95US-0401068.

GenCore version 4.5
Copyright (c) 1993 - 2000 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:23:59 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196F-1

Perfect score: 1 gacattgcgaagtcacagaatla 25

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

N_Geneseq_1101:*

- 1: /SIDS2/gcgcdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDS2/gcgcdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDS2/gcgcdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDS2/gcgcdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDS2/gcgcdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDS2/gcgcdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDS2/gcgcdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDS2/gcgcdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDS2/gcgcdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDS2/gcgcdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDS2/gcgcdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDS2/gcgcdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDS2/gcgcdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDS2/gcgcdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDS2/gcgcdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDS2/gcgcdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDS2/gcgcdata/geneseq/geneseqn/NA1997.DAT:*
- 18: /SIDS2/gcgcdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDS2/gcgcdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDS2/gcgcdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDS2/gcgcdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA901303 E. coli biotin syn
3	25	100.0	1084	10	AA901329 E. coli Bio B gene.
4	25	100.0	1121	7	AA60486 Biotin-biosynthesi
5	25	100.0	3872	15	AA062386 Glutamate receptor
6	18.6	74.4	34980	22	AA68531 EST clone HK26. H
7	17.8	71.2	3083	12	AA011851 F. coli Bio B gene
8	17.6	70.4	512	20	AAV88744 Human gene signatu
9	17.6	70.4	1041	10	AA901327 Human transmembran
10	17.2	68.8	322	16	AA119139
11	17.2	68.8	902	21	AA256756

c 12	17	68.0	17	14	AA056307	BioB DNA primer.
c 13	16.6	66.4	1374	15	AA068434	Maize Hml cDNA. 2
c 14	16.6	66.4	1374	16	AA099463	Maize Hml gene cDN
c 15	16.6	66.4	2157	8	AA070438	Streptococcus pyog
c 16	16.6	66.4	2421	22	AAH20884	H. polymorpha 255
c 17	16.6	66.4	3717	21	AA048934	Arabidopsis thalia
c 18	16.6	66.4	3898	22	AA000666	Human INTERCEPT 28
c 19	16.6	66.4	3925	22	AA000665	Human INTERCEPT 28
c 20	16.6	66.4	3958	22	AA000664	Human INTERCEPT 28
c 21	16.6	66.4	3985	22	AA000663	Human INTERCEPT 28
c 22	16.6	66.4	4018	22	AA000662	Human INTERCEPT 28
c 23	16.6	66.4	4074	22	AA000661	Human INTERCEPT 28
c 24	16.6	66.4	5198	15	AA068433	Maize Hml gene. 2
c 25	16.6	66.4	5198	16	AA099462	Maize Hml genomic
c 26	16.2	64.8	386	21	AA024233	Human secreted pro
c 27	16.2	64.8	1150	21	AA039577	Arabidopsis thalia
c 28	16.2	64.8	1771	22	AA030602	Mouse integrin alp
c 29	16.2	64.8	1771	22	AA010264	Mouse integrin alp
c 30	16.2	64.8	2114	14	AA051236	Plant NADPH cytoch
c 31	16.2	64.8	2697	20	AA078082	A. gossypii G0A2 D
c 32	16.2	64.8	2879	18	AA076318	Xenopus beta-signa
c 33	16.2	64.8	2989	16	AA027939	Human glutamate re
c 34	16.2	64.8	2989	16	AA028000	Human glutamate re
c 35	16.2	64.8	3070	14	AA062694	Human GUR3B recep
c 36	16.2	64.8	32351	21	AA021307	Human low adenosi
c 37	16.2	64.8	32351	21	AA035185	Human low adenosi
c 38	16.2	64.8	40298	21	AA021311	Human adenosine re
c 39	16.2	64.8	40298	21	AA035189	Human adenosine re
c 40	16	64.0	395	21	AA030726	Human colon cancer
c 41	16	64.0	420	18	AA067395	H. pylori surface
c 42	16	64.0	423	19	AA030422	H. pylori cellular
c 43	16	64.0	470	22	AA110226	Probe #159 for gen
c 44	16	64.0	470	22	AA131474	Probe #160 used to
c 45	16	64.0	470	22	AA100164	Probe #155 used to
c 46	16	64.0	477	18	AA068284	H. pylori transmem
c 47	16	64.0	497	22	AA068877	Human cervical can
c 48	16	64.0	537	21	AA036423	Arabidopsis thalia
c 49	16	64.0	889	22	AAH34479	Human colon cancer
c 50	16	64.0	1157	21	AA081657	N. meningitidis pa
c 51	16	64.0	2387	22	AAH15089	Human cDNA sequenc
c 52	16	64.0	2973	21	AA029208	Human topoisomeras
c 53	16	64.0	4812	21	AA013178	Vitellogenin encod
c 54	16	64.0	5994	21	AA049634	Human TUB 5', regio
c 55	16	64.0	5995	18	AA096641	Human TUB locus 5'
c 56	16	64.0	6688	19	AA058288	S. pyogenes SP-26-
c 57	16	64.0	1230025	20	AA051990	Nucleotide sequenc
c 58	15.8	63.2	60	14	AA051028	Fragment of the hu
c 59	15.8	63.2	253	15	AA076727	Human genome fragm
c 60	15.8	63.2	518	22	AA129477	Human tumor relat
c 61	15.8	63.2	567	22	AAH05115	Human cDNA clone (
c 62	15.8	63.2	629	21	AA041591	Arabidopsis thalia
c 63	15.8	63.2	629	21	AA039882	Arabidopsis thalia
c 64	15.8	63.2	834	22	AAH04186	Human cDNA clone (
c 65	15.8	63.2	837	21	AA080144	Fusarium venenatum
c 66	15.8	63.2	1188	19	AA028867	Human integrin mem
c 67	15.8	63.2	1244	22	AAH13961	Human cDNA sequenc
c 68	15.8	63.2	1322	20	AA04402	Human secreted pro
c 69	15.8	63.2	1337	20	AA04345	Human secreted pro
c 70	15.8	63.2	1876	22	AAH15485	Human cDNA sequenc
c 71	15.8	63.2	2718	14	AA051025	Human glutamate re
c 72	15.8	63.2	2911	16	AA027939	Human glutamate re
c 73	15.8	63.2	2946	16	AA027939	Human glutamate re
c 74	15.8	63.2	3220	14	AA054116	Human GUR1B recep
c 75	15.8	63.2	92584	21	AA022288	BAC containing rep
c 76	15.6	62.4	195	22	AAH22655	PPAgamma responsi
c 77	15.6	62.4	291	21	AA031508	Human secreted pro
c 78	15.6	62.4	315	22	AAH22654	PPAgamma responsi
c 79	15.6	62.4	315	22	AAH22656	PPAgamma responsi
c 80	15.6	62.4	350	17	AA010740	CSF2C2, DNA shows H
c 81	15.6	62.4	564	21	AA044922	Arabidopsis thalia
c 82	15.6	62.4	566	22	AA118758	Probe #8691 for ge
c 83	15.6	62.4	566	22	AA118758	Probe #12537 used
c 84	15.6	62.4	633	21	AA280413	Human colon cancer

Phrap Value Range

```

FEATURES
Source
-----
Version: 1.01 qxfo
Location/Qualifiers
1..105207
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12p12-27.2-31.7"
/clone="RP11-18J10"
18..60
/rpt_family="(TA)n"
63..190
/standard_name="G41297"
/db_xref="dbSTS:61467"
complement(494..724)
/rpt_family="MIR"
complement(1141..2037)
/rpt_family="L1MC/D"
2118..2527
/rpt_family="LTR52"
3450..3542
/rpt_family="T-rich"
complement(4520..4976)
/rpt_family="HAL"
complement(5590..5769)
/rpt_family="LTR50"
complement(6173..6317)
/rpt_family="LTR41"
complement(7296..7605)
/rpt_family="AluSp"
complement(8517..8632)
/rpt_family="MLT1A1"
complement(10161..10450)
/rpt_family="AluY"
11042..11157
/rpt_family="L2"
complement(12297..12472)
/rpt_family="MIR"
12499..12537
/rpt_family="L2"
complement(12558..12933)
/rpt_family="MLT1A1"
13472..13730
/standard_name="G41298"
/db_xref="dbSTS:61468"
complement(13981..14015)
/rpt_family="L2"
14109..14172
/rpt_family="MIR"
complement(14173..14808)
/rpt_family="L1MA5"
14809..14822
/rpt_family="MIR"
complement(14919..15071)
/rpt_family="L2"
complement(15133..15254)
/rpt_family="MIR"
complement(15938..16049)
/rpt_family="L1MA3"
18821..19299
/rpt_family="MLT1J1"
complement(19894..20002)
/rpt_family="MIR"
complement(20027..20601)
/rpt_family="MLT2CB"
complement(20652..20823)
/rpt_family="LTR7"
complement(21034..21559)
/rpt_family="MLT1E2"
22407..22453
/rpt_family="(TG)n"

```

Query Match 74.4%; Score 18.6; DB 9; Length 105207;
 Best Local Similarity 84.0%; Pred. No. 88;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacatgtcgaacacacagaatta 25
 ||||| ||| ||||| |||||
 Db 87086 GACATGTGCGCATTTGCACAGTATTA 87062

RESULT 43

CNS01RGD
 LOCUS
 DEFINITION
 Human chromosome 14 DNA sequence BAC R-68M10 of library RPc1-11
 AL157915
 from chromosome 14 of Homo sapiens (human), complete sequence.
 VERSION
 AL157915.3 GI:13940013
 KEYWORDS
 HTG.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 146002)
 Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
 Brothier, P., Catolico, L., Barbe, V., Pellerin, E., Artiguenave, F.,
 Levy, M., Eckenberg, R., Bruls, T., deBardinas, V., Crnaud, C.,
 Gyapay, G., Sautin, W. and Weissbach, J.
 Sequencing of the human chromosome 14
 Unpublished

TITLE

2 (bases 1 to 146002)
 Genoscope.

AUTHORS

Direct Submission
 Submitted (30-APR-2001) Genoscope - Centre National de Sequencage ;
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segefe@genoscope.cns.fr)
 Web : www.genoscope.cns.fr)

JOURNAL

COMMENT

On May 3, 2001 this sequence version replaced gi:7710968.
 ----- Genom Center

Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Segefe@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : C-2177B16 (AC-AL136294)
 Downstream BAC (overlapping the SP6 end) : C-256603 -----
 Summary Statistics
 Assembly program: Phrap; version 2.0
 Quality coverage: 8.93x in Q20 bases; sum-of-configs

Overall quality chart :

Range	Bases
0 - 9	1
10 - 19	42
20 - 29	518
30 - 39	1270
40 - 49	5281
50 - 59	9119
60 - 69	10404
70 - 79	7359
80 - 89	9889
90 - 99	24831
	77308

Percentage of bases with a quality value >= 40 : 95 %.

FEATURES

```

Source
-----
Location/Qualifiers
1..146002
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-68M10"
/clone_id="RPc1-11"
BASE COUNT 47534 a 27851 c 27348 g 43268 t 1 others

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2027E7"
BASE COUNT      32468 a 17525 c 17811 g 32380 t
ORIGIN

```

Query Match	74.4%	Score 18.6	DB 9	Length 100184
Best Local Similarity	84.0%	Pred. No. 88		
Matches	21	Conservative	0	Mismatches
			4	Indels
			0	Gaps
Oy	1	gacattgtgcgaagtcacagaatta	25	
Db	34693	GACAGTCTCTCAATATCACAGATTTA	34717	

RESULT	42
AC011594/c	
LOCUS	AC011594 105207 bp DNA PRI 19-FEB-2000
DEFINITION	Homo sapiens I2P12-27.2-31.7 BAC RP11-18110 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION	AC011594
VERSION	AC011594.8 GI:7007623
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE AUTHORS

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 105207)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (08-OCT-1999) Human Genome Sequencing Center, Department

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect $< 1e-34$) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES. This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Config length: 105207
Phrap values in estimate: 101847
Average error rate (BCM-Phrap estimate): 3.30747e-05
Fraction of Phrap values less than 40 : 0.00530207
Number of N's in consensus edits: 12
Number of N's in consensus : 0
```

Position	Consensus	Changing	edits	Original+Context	Edited+Context
21062	tcacgagcga(n)aacacaccta	tcacgagcga(n)aacacaccta		tcacgagcga(n)aacacaccta	tcacgagcga(n)aacacaccta
21398	ctctcatatc(n)tatattatcc	ctctcatatc(n)tatattatcc		ctctcatatc(n)tatattatcc	ctctcatatc(n)tatattatcc
36393	tcctagagttc(a)nataagtttlc	tcctagagttc(a)nataagtttlc		tcctagagttc(a)nataagtttlc	tcctagagttc(a)nataagtttlc
36340	ctagagttta(n)tatagttttaa	ctagagttta(n)tatagttttaa		ctagagttta(n)tatagttttaa	ctagagttta(n)tatagttttaa
42462	ttgaattatc(n)tttgatatttaa	ttgaattatc(n)tttgatatttaa		ttgaattatc(n)tttgatatttaa	ttgaattatc(n)tttgatatttaa
53293	tcaacgtatg(n)tgactcaagcgc	tcaacgtatg(n)tgactcaagcgc		tcaacgtatg(n)tgactcaagcgc	tcaacgtatg(n)tgactcaagcgc
70017	ttcttccttc(t)ccaatccatc	ttcttccttc(t)ccaatccatc		ttcttccttc(t)ccaatccatc	ttcttccttc(t)ccaatccatc
70019	ctcttccttc(c)atccatcgc	ctcttccttc(c)atccatcgc		ctcttccttc(c)atccatcgc	ctcttccttc(c)atccatcgc
80229	ctattagagc(n)aacattcttc	ctattagagc(n)aacattcttc		ctattagagc(n)aacattcttc	ctattagagc(n)aacattcttc
72637	ttgcagagcag(n)nagaagggaa	ttgcagagcag(n)nagaagggaa		ttgcagagcag(n)nagaagggaa	ttgcagagcag(n)nagaagggaa
72638	tgccagcga(n)taagatggaaac	tgccagcga(n)taagatggaaac		tgccagcga(n)taagatggaaac	tgccagcga(n)taagatggaaac
15142	ggggggagag(n)g)ggaaggggg	ggggggagag(n)g)ggaaggggg		ggggggagag(n)g)ggaaggggg	ggggggagag(n)g)ggaaggggg

----- Distribution of Quality < 40 Bases

#	bases
500	
450	
400	
350	
300	
250	
200	
150	*
100	*
50	*
0	*

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

2 (bases 1 to 154669)
Worley, K.C.
Submitted (01-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:99366078.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HB3M

Center clone name: RP11-83N22

Summary Statistics

Sequencing vector: M13: 108821

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 132678 bases at least Q40

Consensus quality: 143713 bases at least Q30

Consensus quality: 148415 bases at least Q20

Estimated insert size: 148019; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; aarose-tp estimation

Quality coverage: 3.6x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 23 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 19324: contig of 19324 bp in length

19325 19424: gap of unknown length

19425 36072: contig of 16648 bp in length

36073 36172: gap of unknown length

36173 53677: contig of 17505 bp in length

53678 53777: gap of unknown length

53778 68049: contig of 14272 bp in length

68050 68149: gap of unknown length

68150 79650: contig of 11501 bp in length

79651 79750: gap of unknown length

79751 88462: contig of 8712 bp in length

88463 88562: gap of unknown length

88563 96961: contig of 8399 bp in length

96962 97061: gap of unknown length

97062 103046: contig of 5985 bp in length

103047 103146: gap of unknown length

103147 109579: contig of 6833 bp in length

109580 110079: gap of unknown length

110080 116770: contig of 6691 bp in length

116771 116770: gap of unknown length

116771 122217: contig of 5347 bp in length

122218 122317: gap of unknown length

122318 127480: contig of 5163 bp in length

127481 127580: gap of unknown length

127581 130799: contig of 3219 bp in length

130800 130899: gap of unknown length

130900 13840: contig of 2941 bp in length

133841 133940: gap of unknown length

133941 137088: contig of 3148 bp in length

137089 137188: gap of unknown length

137189 140759: contig of 3571 bp in length

140760 140859: gap of unknown length

140860 145882: contig of 4723 bp in length

145883 145882: gap of unknown length

145883 147042: contig of 1360 bp in length

147043 147142: gap of unknown length

147143 148329: contig of 1187 bp in length

148330 148430: gap of unknown length

148430 150556: contig of 2127 bp in length

150557 150657: gap of unknown length

150657 151726: contig of 1070 bp in length

151727 151826: gap of unknown length

151827 153210: contig of 1384 bp in length

153211 153310: gap of unknown length

153311 154669: contig of 1359 bp in length.

FEATURES
source
1. 154669
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-83N22"

BASE COUNT 48347 a 28593 c 27709 g 47760 t 2260 others

ORIGIN

Query Match 75.2% Score 18.8 DB 2: Length 154669;

Best Local Similarity 90.9% Pred. No. 70;

Matches 20: Conservative 0: Mismatches 2: Indels 0: Gaps 0;

Qy 2 acattgtcgcaagtcacagaat 23

Db 85266 ACATTGTGCAAGTCACAGAAAT 85287

RESULT 41

AC008783

LOCUS

DEFINITION

AC008783

AC008783.5

GI:7582549

HTG.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 100184)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

2 (bases 1 to 100184)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-ANG-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 100184)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

On Apr 18, 2000 this sequence version replaced gi:6957669.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.sngc.stanford.edu

Quality: Phrap Quality >=40 99.7% of Sequence;

Estimated total Number of Errors is 1.6.

SMS Content:

SHEC-104571 G58420.

Location/Qualifiers

1. 100184

```

misc_feature /note="similar to EST AW682581 (NID:g7557290)"
3405..3490
/note="similar to EST W23942 (NID:g1300757) zb80d08.r1"
3405..3490
misc_feature /note="similar to EST AU079130 (NID:g6083886)"
3405..3490
/note="similar to EST AW557653 (NID:g7203082)"
3898..3973
/note="similar to EST AU079130 (NID:g6083886)"
3898..3973
/note="similar to EST AW557653 (NID:g7203082)"
3898..3973
/note="similar to EST AW557653 (NID:g7203082)"
3898..3973
/note="similar to EST AW682581 (NID:g7557290)"
3898..3973
/note="similar to EST W23942 (NID:g1300757) zb80d08.r1"
3949..3973
/note="similar to EST A1047587 (NID:g3295874) uh62g06.r1"
5121..5583
/note="similar to EST R88744 (NID:g953571) yp93402.s1"
5807..6239
/note="similar to EST R91001 (NID:g958541) yp93402.r1"
5860..6005
/rpt_family="Alu"
6265..6573
/rpt_family="Alu"
6838..7537
/note="similar to EST AW853284 (NID:g7948801)"
7047..7169
/rpt_family="L1"
7264..7641
/rpt_family="L1"
8853..9846
/rpt_family="L1"
9968..10563
/rpt_family="L1"
10579..11030
/rpt_family="L1"
11031..11148
/rpt_family="Alu"
11149..14541
/rpt_family="L1"
14963..15959
/rpt_family="L1"
15960..16240
/rpt_family="Alu"
16428..16699
/rpt_family="L1"
17098..17414
/note="similar to EST AA917534 (NID:g3057424) o175a06.s1"
17098..17600
/note="similar to EST A1733137 (NID:g5054250) o175a06.x5"
17417..17528
/rpt_family="MIR"
17435..17876
/note="similar to EST A1792280 (NID:g5340496) o175a06.y5"
17829..17936
/note="similar to EST T80329 (NID:g698838) yd07d01.r1"
17846..17884
/note="similar to EST AW682581 (NID:g7557290)"
17846..17936
/note="similar to EST AU079130 (NID:g6083886)"
17846..17936
/note="similar to EST A1119096 (NID:g3519420) ue23c11.y1"
17846..17936
/note="similar to EST AW557653 (NID:g7203082)"
17846..17936
/note="similar to EST A1047587 (NID:g3295874) uh62g06.r1"
17907..17936
/note="similar to EST AL036090 (NID:g5405715)"
18555..18631
/note="similar to EST A1036090 (NID:g5405715)"

```

```

misc_feature 18555..18631
/note="similar to EST T80329 (NID:g698838) yd07d01.r1"
18555..18631
misc_feature /note="similar to EST A1119096 (NID:g3519420) ue23c11.y1"
18555..18631
/note="similar to EST AW557653 (NID:g7203082)"
18555..18631
/note="similar to EST AU079130 (NID:g6083886)"
18555..18631
/note="similar to EST A1047587 (NID:g3295874) uh62g06.r1"
19380..19456
/rpt_family="MIR"
19459..19754
/rpt_family="Alu"
20036..20150
/rpt_family="L1"
20388..20556
/rpt_family="L1"
20909..21415
misc_feature

```

Query Match 75.2% Score 18.8; DB 9; Length 132444;
 Best Local Similarity 90.9% Pred No 70;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 catgtcgcgaqtcacagaatt 24
 ||||| |||||
 Db 122267 CATGTGCACAGTCACAGATT 122288

RESULT 40
 AC076964 154669 bp DNA HTG 05-JAN-2001
 LOCUS Homo sapiens chromosome 3 clone RP11-83N22, WORKING DRAFT SEQUENCE,
 DEFINITION 23 unordered pieces.
 ACCESSION AC076964
 VERSION AC076964.5 GI:12025561
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1 (bases 1 to 154669)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieava,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkelt,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dalborne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,D.,
 Foster,P., Frantz,P., Gabisi,J., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homs,F., Howard,S., Huber,J., Hultky,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
 Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Oviado,R., Pace,A., Payton,B.,
 Ogund,M., Okunolu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Pitmus,E., Pu,L.L.,
 Qullis,M., Ren,Y., Rives,M., Rojas,A., Rojuboakan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,

```

repeat_region 26334..26842
/Note="LIMA2 repeat: matches 532..1055 of consensus"
repeat_region 27345..27453
/Note="AluNb repeat: matches 1..107 of consensus;
Incomplete repeat"
repeat_region 27461..27812
/Note="THEIB repeat: matches 364..1 of consensus"
repeat_region 27813..29193
/Note="THEIB-INTERNAL repeat: matches 1580..198 of
consensus"
repeat_region 29194..29488
/Note="AluSg repeat: matches 300..2 of consensus"
repeat_region 29489..29692
/Note="THEIB-INTERNAL repeat: matches 206..1 of consensus"
repeat_region 29693..30040
/Note="THEIB repeat: matches 364..4 of consensus"
repeat_region 30049..30210
/Note="AluIo repeat: matches 114..282 of consensus;
Incomplete repeat"
repeat_region 30966..32635
/Note="L1 repeat: matches 3679..5382 of consensus"
repeat_region 32498..32671
/Note="L1MC2 repeat: matches 3..178 of consensus"
BASE COUNT 9552 a 5758 c 6033 g 11328 t
ORIGIN
Query Match 75.2%; Score 18.8; DB 9; Length 32671;
Best Local Similarity 90.9%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 acattgtcgaagtcacagaat 23
||||| ||||| ||||| |||||
Db 26394 ACATTGTTGCAAGTCACAGAAT 26373

RESULT 39
AC019226 AC019226 132444 bp DNA PRI 30-SEP-2000
DEFINITION Homo sapiens BAC clone RP11-567F11 from 2, complete sequence.
AC019226 AC019226
VERSION AC019226.4 GI:9858446
KEYWORDS htc.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 132444)
Sullivan,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
AUTHORS 2 (bases 1 to 132444)
Fryman,J., Abbott,A. and Bartniczak,K.
TITLE The sequence of Homo sapiens BAC clone RP11-567F11
JOURNAL Unpublished
REFERENCE
AUTHORS 3 (bases 1 to 132444)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 132444)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 132444)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

```

REFERENCE
AUTHORS 6 (bases 1 to 132444)
Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 20, 2000 this sequence version replaced gi:7574967.
COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H.NH0567F11

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPc1-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-295N18. Actual end of this
clone is at base position 132444 of RP11-567F11.
Location/Qualifiers
1..132444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-567F11"
/clone.lib="RPc1-11"
repeat_region 223..406
/rpt_family="L1"
misc_feature 371..502
/Note="similar to EST AM682581 (NID:97557290)"
misc_feature 371..502
/Note="similar to EST AU079130 (NID:96083886)"
misc_feature 387..502
/Note="similar to EST W23942 (NID:91300757) zB80608.r1"
misc_feature 443..502
/Note="similar to EST AW557653 (NID:97203082)"
repeat_region 1621..1745
/rpt_family="MIR"
repeat_region 2777..3027
/rpt_family="MIR"
misc_feature 3400..3490
/Note="similar to EST A119096 (NID:93519420) uc23c11.y1"
misc_feature 3405..3490

```

* 100463 102354: contig of 1892 bp in length
* 102355 102454: gap of 100 bp
* 102455 104798: contig of 2344 bp in length
* 104799 104898: gap of 100 bp
* 104899 106543: contig of 1645 bp in length
* 106544 106643: gap of 100 bp
* 106644 108168: contig of 1525 bp in length
* 108169 108268: gap of 100 bp
* 108269 110965: contig of 2697 bp in length
* 110966 111065: gap of 100 bp
* 111066 112766: contig of 1701 bp in length
* 112767 112866: gap of 100 bp
* 112867 115404: contig of 2538 bp in length
* 115405 115504: gap of 100 bp
* 115505 117988: contig of 2484 bp in length
* 117989 118088: gap of 100 bp
* 118089 121235: contig of 3147 bp in length
* 121236 121335: gap of 100 bp
* 121336 123563: contig of 2228 bp in length
* 123564 123663: gap of 100 bp
* 123664 125656: contig of 1993 bp in length
* 125657 125756: gap of 100 bp
* 125757 128559: contig of 2803 bp in length
* 128560 128659: gap of 100 bp
* 128660 131968: contig of 3309 bp in length
* 131969 132068: gap of 100 bp

Query Match 76.8%; Score 19.2; DB 2; Length 232409;
Best Local Similarity 87.58; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 acattgcgaagtcacagaatta 25
Db 58572 ACATTGTCGAATACAGATTA 58595
||||||| ||| |||||

RESULT 38
HS2A2A/c 32671 bp DNA PRT 23-NOV-1999
DEFINITION Human DNA sequence from PAC 2A2 on chromosome X.
ACCESSION Z64815
VERSION Z64815.1 GI:1834462
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 32671)
REFERENCE 1
AUTHORS Deadman, R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1997) E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
de Jong P.J., enquiries: http://bacpac.med.buffalo.edu/IMPORTANT:
This sequence is not the entire insert of clone 2A2. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The sequence from clone 2A2 has been finished in more than one
contig. This sequence (2A2A) is separated from the following one
(2A2B) by a gap of 350 bp sized by PCR.
The true left end of clone 2A2 is at 1 in this sequence. 2A2 is
from the human PAC library described in Ioannou A.P. et al Nature
Genet 6, 84-89.

FEATURES

Location/Qualifiers
1..32671
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="X"
/map="X"
/clone_lib="RPI-1"
/clone="RPI-2A2"
659..1020
/note="LI repeat: matches 4769..4409 of consensus"
1005..1354
/note="MSTA repeat: matches 1..391 of consensus"
2554..2641
/note="LI repeat: matches 4423..4336 of consensus"
2641..2759
/note="LI repeat: matches 4785..4904 of consensus"
2751..3498
/note="LIPA9 repeat: matches 147..909 of consensus"
5034..5084
/note="LIPA1 repeat: matches 401..351 of consensus"
5079..5387
/note="MUT28 repeat: matches 1..309 of consensus"
5079..5635
/note="MUT20 repeat: matches 1..546 of consensus"
5642..5979
/note="LMD2 repeat: matches 358..1 of consensus"
6162..6332
/note="LI repeat: matches 4586..4756 of consensus"
6503..7375
/note="LIPA11 repeat: matches 902..1 of consensus"
7239..11641
/note="LI repeat: matches 5390..966 of consensus"
13509..13800
/note="AluX repeat: matches 1..293 of consensus"
13803..14023
/note="LIME1 repeat: matches 860..655 of consensus"
14047..14383
/note="LIME2 repeat: matches 672..334 of consensus"
14385..14679
/note="AluX repeat: matches 1..302 of consensus"
14815..14862
/note="Alu repeat: matches 302..254 of consensus;
incomplete repeat"
14884..15035
/note="LIMB5 repeat: matches 216..63 of consensus"
15227..15574
/note="THE1B repeat: matches 361..5 of consensus"
15988..16442
/note="LIMAV repeat: matches 1017..563 of consensus"
16544..16854
/note="LIMA2 repeat: matches 1..315 of consensus"
17046..17342
/note="AluSc repeat: matches 297..1 of consensus"
17963..18091
/note="LIMB8 repeat: matches 911..778 of consensus"
18106..18563
/note="MUT2CB repeat: matches 1..461 of consensus"
18619..18663
/note="MUT2CB repeat: matches 457..501 of consensus"
18665..20109
/note="MUT2 internal repeat: matches 2..1389 of consensus"
20135..20340
/note="LIMA7 repeat: matches 236..442 of consensus"
20331..20720
/note="LIMS repeat: matches 395..1 of consensus"
20575..23701
/note="LI repeat: matches 5390..2254 of consensus"
23707..24060
/note="THEIC repeat: matches 360..1 of consensus"
24075..24914
/note="LI repeat: matches 2242..1403 of consensus"
24718..25249
/note="MER25 repeat: matches 2101..1583 of consensus"
26154..26302
/note="LI repeat: matches 4200..4059 of consensus"
26302..26366
/note="THEIC repeat: matches 371..305 of consensus"

TITLE
JOURNAL
COMMENT

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teefaye, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 15, 2000 this sequence version replaced gl:8389580.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Clone name: 183_N.5

Center Project name: L7774

NOTE: This is a 'working draft' sequence. It currently
consists of 92 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 315: contig of 315 bp in length
316 415: gap of 100 bp
416 1546: contig of 1131 bp in length
1547 1646: gap of 100 bp
1647 2714: contig of 1068 bp in length
2715 2814: gap of 100 bp
2815 3928: contig of 1114 bp in length
3929 4028: gap of 100 bp
4029 5262: contig of 1234 bp in length
5263 5362: gap of 100 bp
5363 6777: contig of 1415 bp in length
6778 6877: gap of 100 bp
6878 8003: contig of 1126 bp in length
8004 8103: gap of 100 bp
8104 9142: contig of 1039 bp in length
9143 9242: gap of 100 bp
9243 10671: contig of 1429 bp in length
10672 10771: gap of 100 bp
10772 11801: contig of 1030 bp in length
11802 11901: gap of 100 bp
11902 13290: contig of 1389 bp in length
13291 13390: gap of 100 bp
13391 14465: contig of 1075 bp in length
14466 14565: gap of 100 bp
14566 16007: contig of 1442 bp in length
16008 16107: gap of 100 bp
16108 17424: contig of 1317 bp in length
17425 17524: gap of 100 bp
17525 18567: contig of 1043 bp in length
18568 18667: gap of 100 bp
18668 20193: contig of 1526 bp in length
20194 20293: gap of 100 bp
20294 21489: contig of 1196 bp in length
21490 21589: gap of 100 bp
21590 22734: contig of 1145 bp in length
22735 22834: gap of 100 bp
22835 23858: contig of 1024 bp in length
23859 23958: gap of 100 bp
23959 25029: contig of 1071 bp in length
25030 25129: gap of 100 bp
25130 26400: contig of 1271 bp in length

26401 26500: gap of 100 bp
26501 27505: contig of 1005 bp in length
27506 27605: gap of 100 bp
27606 29263: contig of 1658 bp in length
29264 29363: gap of 100 bp
29364 30538: contig of 1175 bp in length
30539 30638: gap of 100 bp
30639 32114: contig of 1476 bp in length
32115 32214: gap of 100 bp
32215 34074: contig of 1860 bp in length
34175 34174: gap of 100 bp
34175 35360: contig of 1186 bp in length
35361 35460: gap of 100 bp
35461 36996: contig of 1536 bp in length
36997 37096: gap of 100 bp
37097 38668: contig of 1572 bp in length
38669 38768: gap of 100 bp
38769 40739: contig of 1971 bp in length
40740 40839: gap of 100 bp
40840 42472: contig of 1633 bp in length
42473 42572: gap of 100 bp
42573 44189: contig of 1617 bp in length
44190 44289: gap of 100 bp
44290 45726: contig of 1437 bp in length
45727 45826: gap of 100 bp
45827 47289: contig of 1463 bp in length
47290 47389: gap of 100 bp
47390 49166: contig of 1777 bp in length
49167 49266: gap of 100 bp
49267 51468: contig of 2202 bp in length
51469 51568: gap of 100 bp
51569 52883: contig of 1315 bp in length
52884 52983: gap of 100 bp
52984 54680: contig of 1637 bp in length
54681 54780: gap of 100 bp
54781 56584: contig of 1804 bp in length
56585 56684: gap of 100 bp
56685 58027: contig of 1343 bp in length
58028 58127: gap of 100 bp
58128 59566: contig of 1439 bp in length
59567 59666: gap of 100 bp
59667 61605: contig of 1939 bp in length
61606 61705: gap of 100 bp
61706 62868: contig of 1163 bp in length
62869 62968: gap of 100 bp
62969 64543: contig of 1575 bp in length
64544 64643: gap of 100 bp
64644 67120: contig of 2477 bp in length
67121 67220: gap of 100 bp
67221 68814: contig of 1594 bp in length
68815 68914: gap of 100 bp
68915 70497: contig of 1583 bp in length
70498 70597: gap of 100 bp
70598 72725: contig of 2128 bp in length
72726 72825: gap of 100 bp
72826 74773: contig of 1948 bp in length
74774 74873: gap of 100 bp
74874 76723: contig of 1850 bp in length
76724 76823: gap of 100 bp
76824 85187: contig of 8364 bp in length
85188 85287: gap of 100 bp
85288 88050: contig of 2763 bp in length
88051 88150: gap of 100 bp
88151 89351: contig of 1201 bp in length
89352 89451: gap of 100 bp
89452 91285: contig of 1834 bp in length
91286 91385: gap of 100 bp
91386 94258: contig of 2873 bp in length
94259 94358: gap of 100 bp
94359 96730: contig of 2372 bp in length
96731 96830: gap of 100 bp
96831 100362: contig of 3532 bp in length
100363 100462: gap of 100 bp


```

83778      aaaaatctn(n)ttttanmn      aaaaatctc(g)ttataagc
83782      tttcnnlt(t)tanmncnc      attccgttt(a)taagctac
83785      ctttttla(n)nmncncat      cctgtttata(a)gctactac
83786      ttttttan(n)nmncncat      cttgtttata(g)ctactaat
83787      ntttttan(n)nmncncat      gttttata(g)c)actcaatt
83788      nttttanm(n)cncncttt      gttttataag(c)actcaatt
83789      ttttanmnc(c)ncnctttt      ttataagc(a)ctcaattgt
83790      tttanmnc(c)ncnctttt      ttataagc(a)ctcaattgt
83791      ttanmnc(n)cnctttgt      ttataagc(a)ctcaattgt
83794      annmnc(na)ctttgttt      taagctac(a)ctcaattgt
83798      nctncaatt(t)gtttttt      taagctac(a)ctcaattgt
83802      nctncaatt(t)gtttttt      taagctac(a)ctcaattgt
83803      caatttgt(n)tttttttt      caatttgt(a)ctcaattgt
83813      ancttgn(c)tttttttt      caatttgt(a)ctcaattgt
83815      tttttttt(t)tmngccca      attttttt(a)ctcaattgt
83816      tttttttt(n)mgcccaaa      attttttt(a)ctcaattgt
83819      tttttttt(n)mgcccaaa      attttttt(a)ctcaattgt
83826      tttttttt(n)mgcccaaa      attttttt(a)ctcaattgt
83828      tttttttt(n)mgcccaaa      attttttt(a)ctcaattgt
83830      tttttttt(n)mgcccaaa      attttttt(a)ctcaattgt

```

Query Match 76.8%; Score 19.2; DB 9; Length 197898;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 acatgtcgaagtcacagaatta 25

Db 192429 ACATTGTCCCAACTACATTAATTA 192452

RESULT 36
 AL450352
 LOCUS Homo sapiens chromosome 1 clone RP11-404H1, *** SEQUENCING IN
 DEFINITION
 ACCESSION AL450352
 VERSION AL450352.17 GI:14575294
 KEYWORDS HTGS; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL 1 (bases 1 to 204154)
 Submitted (04-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14348484.

COMMENT
 Genome Center

Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 Project Information
 Center project name: BA0404H1
 Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 13% of reads Chemistry:
 Dye-terminator Big Dye; 86% of reads
 Consensus quality: 202352 bases at least Q40
 Consensus quality: 202994 bases at least Q30
 Consensus quality: 203311 bases at least Q20
 Insert size: 203854; sum-of-contrigs
 Insert size: 173174; 12.4% error; agarose-fp
 Quality coverage: 7.16x in Q20 bases; sum-of-contrigs Quality
 coverage: 8.48x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 2727: contrig of 2727 bp in length
 * 2728 2827: gap of 100 bp
 * 2828 8226: contrig of 5399 bp in length
 * 8227 8326: gap of 100 bp
 * 8327 11860: contrig of 3534 bp in length
 * 11861 11960: gap of 100 bp
 * 11961 204154: contrig of 192194 bp in length.
 Location/Qualifiers
 1..204154
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-404H1"
 /clone_1id="RPC1-11.2"
 1..2727
 /note="assembly_fragment:03384"
 2828..8226
 /note="assembly_fragment:03717"
 8327..11860
 /note="assembly_fragment:03916"
 11961..204154
 /note="assembly_fragment:03340
 clone_end:17
 vector_side:right"

BASE COUNT 61206 a 36852 c 37666 g 68130 t 300 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 204154;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 acatgtcgaagtcacagaatta 25

Db 130559 ACATTGTCCCAACTACAGATTTA 130582

RESULT 37
 AC025126
 LOCUS Homo sapiens chromosome 18 clone RP11-183N5 map 18, *** SEQUENCING
 DEFINITION
 ACCESSION AC025126
 VERSION AC025126.4 GI:9802846
 KEYWORDS HTGS; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL 1 (bases 1 to 232409)
 Birren, B.; Linton, L.; Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 18, clone RP11-183N5
 Unpublished
 2 (bases 1 to 232409)
 Birren, B.; Linton, L.; Nusbaum, C.; Lander, E.; Abraham, H.; Allen, N.;
 Anderson, S.; Baldwin, J.; Barna, N.; Bastien, V.; Beda, F.;
 Boguski, L.; Boukhalil, B.; Brown, A.; Burkett, G.;
 Campilano, A.; Castle, A.; Choe, Y.; Colangelo, M.; Collins, S.;
 Collumore, A.; Cooke, P.; DeArnell, K.; Dewar, K.; Diaz, J.S.;
 Dodge, S.; Domingo, M.; Doyle, M.; Ferreira, P.; Fitzhugh, W.; Gage, D.;
 Galagan, J.; Gardina, S.; Ginde, S.; Goyette, M.; Graham, L.;
 Grand-Pré, N.; Grant, G.; Hages, B.; Heath, A.; Horton, L.;
 Howland, J.C.; Iliev, I.; Johnson, R.; Jones, C.; Kann, L.; Karatas, A.;
 Klein, J.; Labèque, K.; Lamazares, R.; Landers, T.; Lebeck, J.;
 Levine, R.; Liu, C.; Liu, G.; Locke, K.; Macdonald, P.; Margulis, N.;
 McCarthy, M.; McEwan, P.; McKernan, K.; McPherson, R.;
 Meldrum, J.; Meneus, L.; Mihova, T.; Miranda, C.; Miengo, V.; Morrow, J.;

Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, J., Liu, W., Louie, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Maron, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mel, G., Merscher, S., Metzger, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabat, K., Montgomery, R. T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Ouyang, M., Okunju, G., Oran, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Qullies, M., Ren, Y., Rivers, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wiczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Zuercher, R. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 197898)

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (09-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 197898)

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (02-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 197898)

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (25-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 2, 2001 this sequence version replaced g1:1200439.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT

----- Summary Statistics -----
Contig length: 199229
Phrap values in estimate: 198842
Average error rate (BCM-Phrap estimate): 3.80907e-05
Fraction of Phrap values less than 40 : 0.0105611
Number of consensus changing edits: 263
Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original+Context	Edited+Context
6278	gaacaatat(n)caaacatat	gaacaatat(c)caaacatat
6317	ttagatgta(n)nggactcta	ttagatgta(g)tggaactcta
6318	tagaatgta(n)ggagactctat	tagaatgta(g)tggaactcta
6332	actctatct(n)gcaatgtag	actctatct(n)gcaatgtag
6361	cttggtgac(n)ctggagctac	cttggtgac(n)ctggagctac
11788	tagactatca(n)tagcgagac	tagactatca(a)tagcgagac
13663	ctgcttgct(n)ttcaccttc	ctgcttgct(n)ttcaccttc
15952	aaccttgag(n)caatgatac	aaccttgag(t)caatgatac
41053	aaacacacac(n)caacacata	aaacacacac(a)caacacata
41135	ttcgtaagg(n)ctggcccta	ttcgtaagg(t)ctggcccta
41140	taagtgctg(c)ctctatctc	taagtgctg(t)ctctatctc
41191	caaatataa(n)gtgattccc	caaatataa(t)gtgattccc
41275	attaattag(n)gatgatgtt	attaattag(t)gatgatgtt
41732	aaacctcca(n)tgaccagtt	aaacctcca(c)tgaccagtt
44792	gcatagttta(n)catataatg	gcatagttta(a)catataatg
45080	ttggttacg(n)tagttaaga	ttggttacg(t)tagttaaga
45092	agttaagac(n)ccagggttc	agttaagac(a)ccagggttc
46447	cacatgtga(n)aatgtgcag	cacatgtga(t)aatgtgcag
54259	aaccagatc(n)taattaaat	aaccagatc(c)taattaaat
54264	gactcttag(n)aaatcccaa	gactcttag(a)aaatcccaa
54513	atgtctccal(n)aaatggaga	atgtctccal(a)aaatggaga
54514	tgcttccata(n)aaatggaga	tgcttccata(a)aaatggaga
54522	ttmaatgga(n)gaagtttct	ttmaatgga(a)gaagtttct
61194	atctctccac(n)ncagccccc	atctctccac(c)ncagccccc
61195	tgccccaac(n)cgcccccc	tgccccaac(t)cgcccccc
61384	tgccacata(n)tggtttttt	tgccacata(t)tggtttttt
62062	aaagaaagag(n)catatata	aaagaaagag(a)catatata
70286	atcaagtgag(n)caagtata	atcaagtgag(t)caagtata
83689	catctgtga(n)tcngaagtt	catctgtga(a)tcngaagtt
83692	tcgttgatc(n)gaagttggc	tcgttgatc(a)gaagttggc
83707	gttgctctc(n)ccaatung	gttgctctc(a)ccaatung
83714	ttccccaat(n)ngantcgt	ttccccaat(g)ngantcgt
83715	ttccccaat(n)ngantcgt	ttccccaat(c)ngantcgt
83716	cncacaat(n)ngantcgt	cncacaat(c)ngantcgt
83719	caatnnaga(n)tcgttgng	caatnnaga(a)tcgttgng
83728	aatctgttg(n)gcttgatc	aatctgttg(a)gcttgatc
83735	tgngccttg(n)tcctggac	tgngccttg(a)tcctggac
83742	ttgcttgta(n)caaccagac	ttgcttgta(a)caaccagac
83759	agactctaga(n)aanngaaa	agactctaga(a)aanngaaa
83760	gactctaga(n)aanngaaa	gactctaga(c)aanngaaa
83762	ctctagaaa(n)ngnaaaaat	ctctagaaa(g)ngnaaaaat
83763	ctctagaaa(n)ngnaaaaat	ctctagaaa(t)ngnaaaaat
83765	taganaaan(n)gaaataatc	taganaaan(a)gaaataatc
83770	aanngnaaa(a)atnctntt	aanngnaaa(t)atnctntt
83772	nnngnaaaa(n)atnctntt	nnngnaaaa(a)atnctntt
83774	gnaaaaat(n)ctntttaa	gnaaaaat(c)ctntttaa
83776	gaaaatnct(n)nttttaam	gaaaatnct(c)nttttaam
83777	aaaaatnct(n)nttttaam	aaaaatnct(t)nttttaam

Quality coverage: 6.93 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 74370: contig of 74370 bp in length
* 74371 74470: gap of unknown length
* 74471 85434: contig of 10964 bp in length
* 85435 85534: gap of unknown length
* 85535 104284: contig of 18750 bp in length
* 104285 104384: gap of unknown length
* 104385 122620: contig of 18236 bp in length
* 122621 122720: gap of unknown length
* 122721 180721: contig of 58001 bp in length.

FEATURES

source

1. 180721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-268022"
1. 74370
/note="assembly_name:Contig10"
74471. 85434
/note="assembly_name:Contig6"
85535. 104284
/note="assembly_name:Contig7
clone_end:SpE
vector_side:right"
104385. 122620
/note="assembly_name:Contig8"
122721. 180721
/note="assembly_name:Contig9"

BASE COUNT 57067 a 32188 c 32085 g 36979 t 402 others

COMMENT

Query Match 76.8%; Score 19.2; DB 2; Length 180721;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgcgaagtcacagaatta 25

Db 38357 ACATTGCCAAGTTCATATTA 38380

RESULT 33
AC092895 181121 bp DNA HTG 09-AUG-2001

LOCUS Homo sapiens chromosome 3q clone RP11-268022, WORKING DRAFT
AC092895
SEQUENCE 6 unordered pieces.

AC092895.1 GI:15134379
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaratunga,H.C., Aye,J.R., Banks,T., Barbieri,J.,
Benton,D., Bimagne,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Rubay,C.,
Burch,P., Byrdell,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Guneratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,R., Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loui-Sege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., Meleod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohapatra,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nookentwo,S.,
Ogduh,M., Okwoudu,G., Otaguaye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubenkan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Verba,V., Villalton,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gbbs.R.

Direct Submission

Unpublished
2 (bases 1 to 181121)

Worley,K.C.

Direct Submission

Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HSCP

Center clone name: RP11-268022

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 177994 bases at least Q40

Consensus quality: 178829 bases at least Q30

Consensus quality: 179228 bases at least Q20

Estimated Insert Size: 179615; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-IP estimation

Quality coverage: 6.8x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/cebank/draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 6 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved

1 67642: contig of 67642 bp in length

67643 67742: gap of unknown length

67743 125743: contig of 58001 bp in length

125744 125843: gap of unknown length

125844 144081: contig of 18238 bp in length

144082 144181: gap of unknown length

144182 162930: contig of 18749 bp in length

162931 163030: gap of unknown length

163031 173994: contig of 10964 bp in length

```

* 3074 5896: contig of 2823 bp in length
* 5897 5996: gap of 100 bp
* 5897 8618: contig of 2622 bp in length
* 8619 8718: gap of 100 bp
* 8719 11797: contig of 3079 bp in length
* 11798 11897: gap of 100 bp
* 11898 15760: contig of 3863 bp in length
* 15761 15860: gap of 100 bp
* 15861 21609: contig of 5749 bp in length
* 21610 21709: gap of 100 bp
* 21710 26567: contig of 4858 bp in length
* 26568 26667: gap of 100 bp
* 26668 30165: contig of 3498 bp in length
* 30166 30265: gap of 100 bp
* 30266 35306: contig of 5041 bp in length
* 35307 35406: gap of 100 bp
* 35407 42067: contig of 6661 bp in length
* 42068 42167: gap of 100 bp
* 42168 47823: contig of 5556 bp in length
* 47824 47923: gap of 100 bp
* 47924 55979: contig of 8056 bp in length
* 55980 56079: gap of 100 bp
* 56080 63447: contig of 7368 bp in length
* 63448 63547: gap of 100 bp
* 63548 74052: contig of 10505 bp in length
* 74053 74152: gap of 100 bp
* 74153 88663: contig of 14511 bp in length
* 88664 88763: gap of 100 bp
* 88764 102375: contig of 13612 bp in length
* 102376 102475: gap of 100 bp
* 102476 118310: contig of 15835 bp in length
* 118311 118410: gap of 100 bp
* 118411 132886: contig of 14476 bp in length
* 132887 132986: gap of 100 bp
* 132987 132466: contig of 19480 bp in length
* 132467 152567: gap of 100 bp
* 152567 173532: contig of 20966 bp in length.

```

FEATURES

```

source
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="2"
  /map="2"
  /clone_lib="RP11-362113"
  /clone_lib="RP11-362113"
  1. 1051
misc_feature
  /note="assembly_fragment"
  1152..2973
misc_feature
  /note="assembly_fragment"
  3074..5896
misc_feature
  /note="assembly_fragment"
  5997..8618
misc_feature
  /note="assembly_fragment"
  8719..11797
misc_feature
  /note="assembly_fragment"
  11898..15760
misc_feature
  /note="assembly_fragment"
  15861..21609
misc_feature
  /note="assembly_fragment"
  21710..26567
misc_feature
  /note="assembly_fragment"
  26668..30165
misc_feature
  /note="assembly_fragment"
  30266..35306
misc_feature
  /note="assembly_fragment"
  35407..42067
misc_feature
  /note="assembly_fragment"
  42168..47823
misc_feature
  /note="assembly_fragment"
  47924..55979
  /note="assembly_fragment"
  clone_end:T7
  vector_side:right"

```

```

misc_feature 56080..63447
              /note="assembly_fragment"
              63548..74052
              /note="assembly_fragment"
              74153..88663
              /note="assembly_fragment"
              88764..102375
              /note="assembly_fragment"
              102476..118310
              /note="assembly_fragment"
              118411..132886
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:left"
misc_feature 132987..152466
              /note="assembly_fragment"
              152567..173532
              /note="assembly_fragment"
BASE COUNT 56172 a 28809 c 28612 g 57932 t 2007 others
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 2; Length 173532;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 gacatgtcgcagtcacagaatt 24
Db 157729 GACATTGCTCAATGACGAAATT 157752

```

RESULT 32

AC073466

LOCUS AC073466 180721 bp DNA HTG 29-AUG-2000

DEFINITION Homo sapiens chromosome 3 clone RP11-268022, WORKING DRAFT

SEQUENCE, 5 unordered pieces.

AC073466

AC073466.3 GI:9937810

VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 180721)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 180721)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (18-JUN-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Aug 29, 2000 this sequence version replaced gi:8810512.

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0268022
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178101 bases at least Q40
Consensus quality: 178984 bases at least Q30
Insert size: 188000; agarose-fp
Insert size: 180321; sum-of-contigs
Quality coverage: 6.61 in Q20 bases; agarose-fp

```

```

/misc_feature /clone="RP11-764C7"
/misc_feature /note="assembly-fragment"
1. .1173
/misc_feature 1274. .2519
/misc_feature /note="assembly-fragment"
2620. .4794
/misc_feature /note="assembly-fragment"
4895. .6532
/misc_feature /note="assembly-fragment"
6633. .7724
/misc_feature /note="assembly-fragment"
7825. .9453
/misc_feature /note="assembly-fragment"
9554. .10857
/misc_feature /note="assembly-fragment"
10958. .12737
/misc_feature /note="assembly-fragment"
12838. .15810
/misc_feature /note="assembly-fragment"
15911. .18882
/misc_feature /note="assembly-fragment"
18983. .21649
/misc_feature /note="assembly-fragment"
21750. .24680
/misc_feature /note="assembly-fragment"
24781. .28513
/misc_feature /note="assembly-fragment"
28614. .31816
/misc_feature /note="assembly-fragment"
31917. .35271
/misc_feature /note="assembly-fragment"
35372. .38395
/misc_feature /note="assembly-fragment"
clone_end:T7
vector_side:right"
38496. .41955
/misc_feature /note="assembly-fragment"
42056. .45874
/misc_feature /note="assembly-fragment"
45975. .50693
/misc_feature /note="assembly-fragment"
50796. .55055
/misc_feature /note="assembly-fragment"
55156. .60637
/misc_feature /note="assembly-fragment"
60738. .65254
/misc_feature /note="assembly-fragment"
65355. .70460
/misc_feature /note="assembly-fragment"
clone_end:SP6
vector_side:left"
70561. .78592
/misc_feature /note="assembly-fragment"
78693. .86070
/misc_feature /note="assembly-fragment"
86171. .93277
/misc_feature /note="assembly-fragment"
93378. .101958
/misc_feature /note="assembly-fragment"
102059. .113198
/misc_feature /note="assembly-fragment"

```

```

Query Match 76.8%; Score 19.2; DB 2; Length 167292;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 2 acattgtcgaagtcacagaatta 25
||||| ||||| |||||
Db 134143 ACATGTCGCAAGTACATTAATTA 134166

```

RESULT 31

```

AC040890
LOCUS 173532 bp DNA HTG 11-APR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-362113 map 2, WORKING DRAFT
AC040890
AC040890.1 GI:7533996
VERSION AC040890.1
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 173532)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 2, clone RP11-362113
Unpublished
2 (bases 1 to 173532)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, M., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kamm, L., Karatas, A.,
Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Severy, P., Spencer, B.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 362-11.13

```

```

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 164477 bases at least Q40
Consensus quality: 168698 bases at least Q30
Consensus quality: 170320 bases at least Q20
Insert size: 171532; sum-of-coverage: 4.8 in Q20 bases; sum-of-coverage
Quality coverage: 4.8 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1051: contig of 1051 bp in length
* 1052 1151: gap of 100 bp
* 1152 2973: contig of 1822 bp in length
* 2974 3073: gap of 100 bp

```

Oy 1 gacattgcgaagtcacagaatt 24
 ||||| ||||| ||||| ||||| |||||
 Db 58295 GACATTGCTCAATGACAGAAATT 58318

RESULT 30
 AC068365
 LOCUS
 DEFINITION Homo sapiens chromosome 3 clone RP11-764C7 map 3, WORKING DRAFT
 AC068365
 AC068365.2 GI:8705056
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 167292)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barina,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Cooke,P., Deaellano,K., Dewar,K., Diaz,J.S.,
 Duggan,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hages,B., Heathford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,R., Lacroque,K., Lamazares,R., Landers,T., Lebecky,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McEwen,K., McEwen,R.,
 Melchior,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Piere,N.,
 Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testiere,S., Theodore,J., Tirrell,A., Travers,M., Tiggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainou,R., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 25, 2000 this sequence version replaced gi:7677736.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L10186
 Center clone name: 764.C.7

Summary Statistics
 Sequencing vector: M13; M77815, 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151203 bases at least Q40
 Consensus quality: 159094 bases at least Q30
 Consensus quality: 161993 bases at least Q20
 Insert size: 170000; agarose-fp
 Insert size: 164292; sum-of-ctrls
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 3.8 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces

FEATURES

Source

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1174 1273: contig of 1173 bp in length
 1274 2519: contig of 1246 bp in length
 2520 2619: gap of 100 bp
 2620 4794: contig of 2175 bp in length
 4795 4894: gap of 100 bp
 4895 6532: contig of 1638 bp in length
 6533 6632: gap of 100 bp
 6633 7724: contig of 1092 bp in length
 7725 7824: gap of 100 bp
 7825 9453: contig of 1629 bp in length
 9454 9553: gap of 100 bp
 9554 10857: contig of 1304 bp in length
 10858 10957: gap of 100 bp
 10958 12737: contig of 1780 bp in length
 12738 12837: gap of 100 bp
 12838 15810: contig of 2973 bp in length
 15811 15910: gap of 100 bp
 15911 18882: contig of 2972 bp in length
 18883 18982: gap of 100 bp
 18983 21649: contig of 2867 bp in length
 21650 21749: gap of 100 bp
 21750 24680: contig of 2931 bp in length
 24681 24780: gap of 100 bp
 24781 28513: contig of 3733 bp in length
 28514 28613: gap of 100 bp
 28614 31816: contig of 3203 bp in length
 31817 31916: gap of 100 bp
 31917 35271: contig of 3355 bp in length
 35272 35371: gap of 100 bp
 35372 38395: contig of 3024 bp in length
 38396 38495: gap of 100 bp
 38496 41955: contig of 3460 bp in length
 41956 42055: gap of 100 bp
 42056 45874: contig of 3819 bp in length
 45875 45974: gap of 100 bp
 45975 50695: contig of 4721 bp in length
 50696 50795: gap of 100 bp
 50796 55055: contig of 4260 bp in length
 55056 55155: gap of 100 bp
 55156 60637: contig of 5482 bp in length
 60638 60737: gap of 100 bp
 60738 65254: contig of 4517 bp in length
 65255 65354: gap of 100 bp
 65355 70460: contig of 5106 bp in length
 70461 70560: gap of 100 bp
 70561 78592: contig of 8032 bp in length
 78593 78692: gap of 100 bp
 78693 86070: contig of 7378 bp in length
 86071 86170: gap of 100 bp
 86171 93277: contig of 7107 bp in length
 93278 93377: gap of 100 bp
 93378 101958: contig of 8581 bp in length
 101959 102058: gap of 100 bp
 102059 113198: contig of 11140 bp in length
 113199 113298: gap of 100 bp
 113299 126286: contig of 12988 bp in length
 126287 126386: gap of 100 bp
 126387 140169: contig of 13783 bp in length
 140170 140260: gap of 100 bp
 140261 167292: contig of 27023 bp in length.

Location/Qualifiers
 1. 167292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3"

Qulies, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R.,
 Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Walbach, M.,
 Wallington, S., Weinstein, G., Weinstein, I., R., Williamson, A.,
 Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and
 Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 149971)
 Worley, K.C.
 Direct Submission
 Submitted (22-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 25, 2000 this sequence version replaced gi:8705281.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HBAE
 Center clone name: RP11-79A14
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 130792 bases at least Q40
 Consensus quality: 142171 bases at least Q30
 Consensus quality: 145357 bases at least Q20
 Estimated insert size: 145483; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 46806: contig of 46806 bp in length
 46807 46906: gap of unknown length
 46907 79857: contig of 32951 bp in length
 79858 79957: gap of unknown length
 79958 99924: contig of 19967 bp in length
 99925 100024: gap of unknown length
 100025 116534: contig of 16510 bp in length
 116535 116634: gap of unknown length
 116635 122163: contig of 5529 bp in length
 122164 122263: gap of unknown length
 122264 128396: contig of 6133 bp in length
 128397 128496: gap of unknown length
 128497 134588: contig of 6092 bp in length
 134589 134688: gap of unknown length
 134689 138579: contig of 3891 bp in length
 138580 138679: gap of unknown length
 138680 141815: contig of 3136 bp in length
 141816 141915: gap of unknown length
 141916 144532: contig of 2617 bp in length
 144533 144632: gap of unknown length
 144633 146694: contig of 2062 bp in length
 146695 146794: gap of unknown length
 146795 148808: contig of 2014 bp in length
 148809 148908: gap of unknown length
 148909 149711: contig of 1063 bp in length.

FEATURES

source

1. 149971
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-79A14"

BASE COUNT 44656 a 29149 c 28710 g 46204 t 1252 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 149971;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acatgtcgcagtcacagaatta 25
 ||||| | ||||| |||||
 Db 24626 ACATTGCTCTAGTCACATMAATTA 24603

RESULT 29

LOCUS

AL161796

DEFINITION

Homo sapiens chromosome 1 clone RP11-487023 map q31.1-31.3, ***

ACCESSION

AL161796

KEYWORDS

HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

Glithero, R.

Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Jun 8, 2001 this sequence version replaced gi:14269907.

COMMENT

----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA487023
 ----- Summary Statistics
 Assembly program: XCAP4; version 4.5
 Sequencing vector: Plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 70% of reads
 Chemistry: Dye-terminator ET-amersham; 29% of reads
 Consensus quality: 159908 bases at least Q40
 Consensus quality: 159908 bases at least Q30
 Consensus quality: 160081 bases at least Q20
 Insert size: 160431; sum-of-contigs
 Insert size: 105263; 71.7% error; agarose-fp
 Quality coverage: 6.19x in Q20 bases; sum-of-contigs Quality
 coverage: 9.79x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

1. 160431
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q31.1-31.3"
 /clone="RP11-487023"
 /clone_lib="RPC1-11.2"
 1. 160431
 misc_feature /note="assembly_fragment:00314"
 BASE COUNT 53928 a 26336 c 26806 g 53361 t
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 160431;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0;


```

repeat_region      /note="MIR repeat: matches 49, .134 of consensus"
                    47306, .47519
                    /note="MIR repeat: matches 7, .246 of consensus"
repeat_region      49116, .49229
                    /note="MIR repeat: matches 137, .262 of consensus"
repeat_region      49230, .49558
                    /note="Aluub repeat: matches 12, .311 of consensus"
repeat_region      49559, .49724

```

```

Query Match      76.8%; Score 19.2; DB 9; Length 107172;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      2 acattgtcgcagtcacagaatla 25
Db      1676 ACATTGTCGCAGTCACAGATCA 1653

```

```

RESULT 27
AC055761
LOCUS      AC055761 146585 bp DNA HTG 06-JUL-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-73L6, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION AC055761
VERSION   AC055761.5 GI:8699867
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens

```

```

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 146585)
            Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
            Bodola,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
            Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondajewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichner,O., Liu,J., Liu,W., Logan,O., Lozada,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabnah,M.,
            Wallington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.

```

```

TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 146585)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (18-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Jun 25, 2000 this sequence version replaced gi:8170780.

```

```

COMMENT
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
-----
Center project name: HBAF
Center clone name: RP11-73L6
-----
Summary Statistics
-----
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 132601 bases at least Q40
Consensus quality: 140586 bases at least Q30

```

```

Consensus quality: 144168 bases at least Q20
Estimated insert size: 144485; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-ff estimation
Quality coverage: 4x in Q20 bases; sum-of-contrigs estimation

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence; it currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 49434: contig of 49434 bp in length
* 49435 49534: gap of unknown length
* 49535 76020: contig of 26486 bp in length
* 76021 76120: gap of unknown length
* 76121 106128: contig of 30008 bp in length
* 106129 106228: gap of unknown length
* 106229 130157: gap of 23929 bp in length
* 130158 130257: gap of unknown length
* 130258 139669: contig of 9412 bp in length
* 139670 139769: gap of unknown length
* 139770 145465: contig of 5696 bp in length
* 145466 145565: gap of unknown length
* 145566 145585: contig of 1020 bp in length.

```

```

FEATURES
source
1..146585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-73L6"

```

```

BASE COUNT 43675 a 28441 c 28398 g 45447 t 624 others
ORIGIN

```

```

Query Match      76.8%; Score 19.2; DB 2; Length 146585;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      2 acattgtcgcagtcacagaatla 25
Db      146028 ACATTGTCGTCAGTCACATTAATTA 146051

```

```

RESULT 28
AC063937
LOCUS      AC063937 149971 bp DNA HTG 25-JUL-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-79A14, WORKING DRAFT SEQUENCE,
            13 unordered pieces.
ACCESSION AC063937
VERSION   AC063937.8 GI:9438590
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens

```

```

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 149971)
            Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
            Bodola,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
            Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondajewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichner,O., Liu,J., Liu,W., Logan,O., Lozada,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,

```

```
/map="g23"
/clone="RP4-737M10"
/clone_lib="RPct-4"
83..261
repeat_region /note="MIR repeat: matches 73..251 of consensus"
217..278
repeat_region /note="L2 repeat: matches 2649..2708 of consensus"
629..755
repeat_region /note="L2 repeat: matches 2624..2750 of consensus"
1151..1454
repeat_region /note="L2 repeat: matches 2..305 of consensus"
1457..1496
repeat_region /note="Alu repeat: matches 2..305 of consensus"
2387..2853
repeat_region /note="20 copies 2 mer tt 98% conserved"
3497..3931
misc_feature /note="MLT2CH repeat: matches 1..440 of consensus"
3497..3931
/misc_feature /note="match: GSS B39350"
3697..4209
repeat_region /note="MER73 repeat: matches 126..636 of consensus"
5095..5149
repeat_region /note="Alu repeat: matches 242..302 of consensus"
5714..6017
repeat_region /note="Alu repeat: matches 1..304 of consensus"
6511..7165
repeat_region /note="L1PA2 repeat: matches 5487..6142 of consensus"
7229..7932
repeat_region /note="L1PA1 repeat: matches 5438..6147 of consensus"
8052..8346
repeat_region /note="L1M1 repeat: matches 1..298 of consensus"
8421..8612
repeat_region /note="AluSP repeat: matches 51..256 of consensus"
9066..9307
repeat_region /note="MIR repeat: matches 7..262 of consensus"
9865..9918
repeat_region /note="MIR repeat: matches 7..262 of consensus"
9920..10106
repeat_region /note="27 copies 2 mer at 82% conserved"
10155..10717
repeat_region /note="L1MB6 repeat: matches 5782..5965 of consensus"
10705..11191
repeat_region /note="L1M1 repeat: matches 5180..5741 of consensus"
11192..11590
repeat_region /note="L1MB6 repeat: matches 5293..5791 of consensus"
11591..12006
repeat_region /note="L1PA5 repeat: matches 5742..6141 of consensus"
12007..12121
repeat_region /note="L1MB6 repeat: matches 4868..5293 of consensus"
12122..12744
repeat_region /note="FLAM_A repeat: matches 1..127 of consensus"
12699..12925
repeat_region /note="L1MB6 repeat: matches 4266..4868 of consensus"
12913..13353
repeat_region /note="L1 repeat: matches 3741..3964 of consensus"
13359..13675
repeat_region /note="L1 repeat: matches 2997..3429 of consensus"
13844..14150
repeat_region /note="L1M4 repeat: matches 820..1136 of consensus"
14214..14968
repeat_region /note="Alu repeat: matches 1..309 of consensus"
15355..15689
repeat_region /note="L1M4 repeat: matches -11..771 of consensus"
16897..17211
repeat_region /note="match: EST AA814508"
19023..19373
repeat_region /note="Alu repeat: matches 1..306 of consensus"
19598..19850
repeat_region /note="L1R16C repeat: matches 3..385 of consensus"
22634..22735
repeat_region /note="MIR repeat: matches 2..262 of consensus"
23125..23266
repeat_region /note="MIR repeat: matches 90..188 of consensus"
23298..23753
repeat_region /note="MIR repeat: matches 35..180 of consensus"
/note="MIR repeat: matches 35..505 of consensus"

misc_feature complement(23577..23810)
/note="match: STS LA7507"
24138..24348
repeat_region /note="MER20 repeat: matches 3..218 of consensus"
24418..24449
repeat_region /note="MER20 repeat: matches 3..218 of consensus"
25240..25363
repeat_region /note="16 copies 2 mer tt 88% conserved"
25517..25709
repeat_region /note="L2 repeat: matches 2583..2710 of consensus"
25936..26292
repeat_region /note="MER8A repeat: matches 1..188 of consensus"
26464..26646
repeat_region /note="L2 repeat: matches 2367..2727 of consensus"
27233..27384
repeat_region /note="AluSC repeat: matches 122..304 of consensus"
27637..27686
repeat_region /note="MIR repeat: matches 33..211 of consensus"
27774..27864
repeat_region /note="25 copies 2 mer ac 82% conserved"
28116..28419
repeat_region /note="FLAM_A repeat: matches 26..116 of consensus"
28819..28878
repeat_region /note="AluSX repeat: matches 1..311 of consensus"
29192..29378
repeat_region /note="MIR repeat: matches 84..139 of consensus"
29836..31198
repeat_region /note="L1R33 repeat: matches 14..197 of consensus"
31534..31700
repeat_region /note="L1MB2 repeat: matches 4781..6166 of consensus"
32769..32852
repeat_region /note="MIR repeat: matches 66..253 of consensus"
34285..34480
repeat_region /note="MER5B repeat: matches 53..135 of consensus"
34506..34557
repeat_region /note="MER91A repeat: matches 1..195 of consensus"
34701..34775
repeat_region /note="MER5B repeat: matches 1..51 of consensus"
35221..35587
repeat_region /note="MIR repeat: matches 70..142 of consensus"
35984..36131
repeat_region /note="MIR repeat: matches 76..144 of consensus"
36875..36940
repeat_region /note="MER5A repeat: matches 47..189 of consensus"
36989..37342
repeat_region /note="MIR repeat: matches 88..157 of consensus"
37012..37248
repeat_region /note="match: GSS B35055"
37340..37340
repeat_region /note="MIR repeat: matches 1..260 of consensus"
40439..40566
repeat_region /note="match: STS LA1106"
40942..41316
repeat_region /note="L2 repeat: matches 2583..2700 of consensus"
42000..42180
repeat_region /note="L2 repeat: matches 1844..2263 of consensus"
43015..43321
repeat_region /note="Alu repeat: matches 133..311 of consensus"
43339..43646
repeat_region /note="AluSQ repeat: matches 1..305 of consensus"
43831..44112
repeat_region /note="L2 repeat: matches 1631..1942 of consensus"
44165..44425
repeat_region /note="AluSX repeat: matches 1..302 of consensus"
44569..44652
repeat_region /note="L2 repeat: matches 2252..2538 of consensus"
44743..44989
repeat_region /note="L2 repeat: matches 2591..2667 of consensus"
46105..46268
repeat_region /note="MIR repeat: matches 3..248 of consensus"
46155..46506
repeat_region /note="MIR repeat: matches 94..257 of consensus"
47163..47247
repeat_region /note="match: STS AL021628"
```

/note="coded for by C. elegans cDNA yk532c12.5; coded for by C. elegans cDNA yk532c12.3"

/codon_start=1

/product="Hypothetical protein C39D10.5"

/protein_id="AAK39210.1"

/db_xref="GI:13775392"

/translation="MTEKEVELPCGOAVGVFPNIMAPYTELMELAFRDVPGITVA AITLGASDPAVNGTPLSQAAPPDSTQPEVSSTOPDPSAQLTGSRRKSSSPKKWDE LHKRSNHNKTESSEPTDKHCCMFOMAHQLOAKYGVPTPYBNCJTASAPRLMIQYGL PVPVNPQEFNNIGIPRECVKLCGNTEFMPYLK"

complement(22885..24341)

/gene="C39D10.6"

complement(101n(22885..23031,23081..23154,23202..23376,23890..23980,24022..24125,24210..24341))

/gene="C39D10.6"

/codon_start=1

/evidence=not_experimental

/product="Hypothetical protein C39D10.6"

/protein_id="AAK39208.1"

/db_xref="GI:13775392"

/translation="MRGVSEYFEDFPDFHMAATANRVPFQNDKKYKLCGQLHK KGRIVAILNITLAVNIFSEFMRSTVFVTCMSTAPVIFGSLGVMKRLYL VYLFPOSILKVPNSLVLPOVISLISITVLFELISANGASMNVIDIARNVNDPTAS VQEKIDADLALFTVLEFLCLGGLIQATFAEYITSFHNFILKRENSSEFHEYSNS AFGSEATPPPPPPYPAQVSA"

complement(24901..24972)

/note="codon recognized: GAG"

/product="tRNA-Glu"

33178..33249

/note="codon recognized: GAG"

/product="tRNA-Glu"

33178..33249

/note="codon recognized: GAG"

/product="tRNA-Glu"

33557..33712

/gene="C39D10.1"

33557..33712

/gene="C39D10.1"

33557..33712

/gene="C39D10.1"

/codon_start=1

/evidence=not_experimental

/product="Hypothetical protein C39D10.1"

/protein_id="AAK39205.1"

/db_xref="GI:13775389"

/translation="MTTGKDDVDYTTTQNMIFTSPPVNVLESCSADRYTAASAPADEK KISSNOT"

complement(34615..40431)

/gene="C39D10.7"

complement(101n(34615..35022,35071..35167,35214..35223,35369..35483,35581..35736..36757,37399..37659,37706..37951,38129..38551,38623..38806,38852..39021,39075..39179,39235..39339,40376..40431))

/gene="C39D10.7"

/note="similar to C. elegans protein B0280.5, R02F2.4, and C07G2.1; weakly similar to maltose operon periplasmic protein precursor (SP:MAM_SALT, P26478); coded for by C. elegans cDNA yk7c10.3; coded for by C. elegans cDNA yk7c10.5; coded for by C. elegans cDNA yk9b9.5; coded for by C. elegans cDNA yk10d6.3; coded for by C. elegans cDNA yk10d6.5; coded for by C. elegans cDNA yk14c12.3; coded for by C. elegans cDNA yk16h11.5; coded for by C. elegans cDNA yk19h4.3; coded for by C. elegans cDNA yk19h4.5; coded for by C. elegans cDNA yk20a1.3; coded for by C. elegans cDNA yk27d10.3; coded for by C. elegans cDNA yk29b2.3; coded for by C. elegans cDNA yk29b2.5; coded for by C. elegans cDNA yk30a6.5; coded for by C. elegans cDNA yk31b9.3; coded for by C. elegans cDNA yk31b9.5; coded for by C. elegans cDNA yk31d5.3; coded for by C. elegans cDNA yk31d5.5; coded for by C. elegans cDNA yk32d12.5; coded for by C. elegans cDNA yk33a2.3; coded for by C. elegans cDNA yk33a2.5; coded for by C. elegans cDNA yk37e1.3; coded for by C. elegans cDNA yk37e1.5; coded for by C. elegans cDNA yk42a2.3; coded for by C. elegans cDNA yk42a2.5; coded for by C. elegans cDNA yk42c7.3; coded for by C. elegans cDNA yk42e2.5; coded for by C. elegans cDNA yk43b4.3; coded for by C. elegans cDNA

Query Match 76.8%; Score 19.2; DB 3; Length 40897;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 gacatgtgcgaagtcacagaatt 24
 ||||| ||||| ||| |||||
 Db 5770 GACATTGTCGCAATTCAATAAATT 5747

RESULT 26
 HS737M10/C
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

HS737M10 107172 bp DNA PRI 23-NOV-1999
 Human DNA sequence from clone 737M10 on chromosome Xq23. Contains ESTs, an STS and GSSS, complete sequence.
 AL031388
 AL031388.1 GI:3980442
 HTG.
 human.
 Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Submitted (28-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Direct Submission
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Dec 8, 1998 this sequence version was replaced gi:3646111.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 IMPORTANT: This sequence is not the entire insert of clone 737M10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

COMMENT

The true right end of clone 26905 (AC005191) is at 100 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/ChrX
 737M10 is from the library RpiI constructed at the Roswell Park Cancer Institute by the group of Peter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: PCYPA2.

FEATURES

source
 1..107172
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"

misc_feature	/note="assembly_fragment" 32421..40592				
misc_feature	/note="assembly_fragment" 41093..49538				
misc_feature	/note="assembly_fragment" 49639..57982				
misc_feature	/note="assembly_fragment" 58083..73752				
misc_feature	/note="assembly_fragment" 73853..90559				
misc_feature	/note="assembly_fragment" 90660..108100				
misc_feature	/note="assembly_fragment" 108201..128050				
misc_feature	/note="assembly_fragment" 128151..172825				
misc_feature	/note="assembly_fragment" 172926..228434				
misc_feature	/note="assembly_fragment"				
BASE COUNT	73186 a 43737 c 41722 g 67487 t	2302	others		
ORIGIN					
Query Match	79.2%; Score 19.8; DB 2; Length 228434;				
Best Local Similarity	91.3%; Pred. No. 21;				
Matches	21; Conservative	0; Mismatches	2; Indels	0; Gaps	0
QY	2 acattgtcgcaagtcacagaatt 24				
Db	19903 ACATTGTGACAGTGACAGATT 19881				
RESULT 25					
CELC39D10/c					
LOCUS	CELC39D10 40897 bp DNA	INV	24-APR-2001		
DEFINITION	Caenorhabditis elegans cosmid C39D10, complete sequence.				
ACCESSION	U036678				
VERSION	U036678.1 GI:1049439				
KEYWORDS	HTG.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans.				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida;				
	Rhabdilloidea; Rhabdillidae; Pelodolerinae; Caenorhabditis.				
	1 (bases 1 to 40897)				
REFERENCE	The C. elegans Genome Sequencing Consortium, Washington University				
AUTHORS	Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,				
	Hinxton, U.K., C.				
	Genome sequence of the nematode C. elegans: a platform for				
	investigating biology. The C. elegans Sequencing Consortium				
	Science 282 (5396), 2012-2018 (1998)				
TITLE					
JOURNAL	2 (bases 1 to 40897)				
REFERENCE	Minx, M.				
AUTHORS	The sequence of C. elegans cosmid C39D10				
JOURNAL	unpublished				
REFERENCE	3 (bases 1 to 40897)				
AUTHORS	Waterson, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-Oct-1995) Robert Waterson				
REFERENCE	4 (bases 1 to 40897)				
AUTHORS	Waterson, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-APR-2001) Department of Genetics, Washington				
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
AUTHORS	Submitted by:				
JOURNAL					
TITLE					
COMMENT					

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

WARNING: These data have only had automated annotation and have not yet been subjected to manual review of that annotation. We will be manually reviewing this information as quickly as possible and at that time this GenBank record will be updated and this warning removed.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation)

FEATURES

source
1. .4089/
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/clone="CEIC39D10"
.6465. .7026
/gene="C39D10.4"
join(6465. .6656,6857. .6940,7000. .7026)
/gene="C39D10.4"
/note="5' exon lies in array of tandem repeats (9 copies of a 24mer) and similar to Icc nucleation protein; coded for by C. elegans cDNA yk93f12.3; coded for by C. elegans cDNA yk93f12.5"
/codon_start=1
/protein_id="AAK39207.1"
/db_xref="GI:13775391"
/translation="MRGSLVTRVSMWLVTKYSMWLVTVKYSMWLVTKYSMWLVTKYSMWLVTKYSMWLVTKLTHAVFASNRDHIVSMFEFGVLRRHLVPIIRDL"
10962. .13972
/gene="C39D10.3"
join(10962. .11035,11089. .11182,11275. .11355,11990. .12085,12129. .12230,13043. .13153,13203. .13280,13330. .13401,13715. .13845,13903. .13972)
/gene="C39D10.3"
/note="similar to B. subtilis glycerol-3-phosphate cytidyllyltransferase (SP:TACD_BACSU, p27623); coded for by C. elegans cDNA yk74ze10.5; coded for by C. elegans cDNA yk74ze10.3"
/codon_start=1
/protein_id="AAK39211.1"
/db_xref="GI:13775395"
/translation="MPKAIVOGLTIPDGKKRKYITDCGCFDFHIFANARLLMPPAKOYGRILVIGSDDELNNGLIPLTDEDERFLISATRWVDIELAGRLMLVTKKHHMTDS ILDEFSAITLPALDPLSNRPVSYSLFQNTFAFVYIGRKPKYTDRKVUYSGADTF HAGHSLEPAKADLGDIYLIVGICDDDVNNEKGITFFPMNLLETRLTSSIKTYDEVF VGPATVSDFEYNLIKRASKIAVSTSETHPRRFADCTYHNRIIEVPDYDATCEEILLERTSRKIASSDSDKEIATAMFNAPLINNH"
15307. .16214
/gene="C39D10.2"
join(15307. .15585,15791. .16034,16111. .16214)
/gene="C39D10.2"
/codon_start=1
/evidence="not_experimental"
/product="Hypothetical protein C39D10.2"
/protein_id="AAK39206.1"
/db_xref="GI:13775390"
/translation="HRQLDTLYLMNEQCKQRCFCDEMRRKSRLPRLSINRRLLPYTCQGLIDKSCQHSTFTLSLTFELFKFPEVYSIFTEGGSINKNOQISGA SDISNFQQLITFEFLAVLCVEAKRNQCDDLVCORMTTEECESIMHKOCRGKRSOPSTPEILEEVIANBLTGIEQAIVELLELAARYPPRRILEXYIH"
complement(17230. .18984)
/gene="C39D10.5"
complement(join(17230. .17244,17951. .18424,18892. .18984))
/gene="C39D10.5"


```

* 1166 3026: contig of 1861 bp in length
* 3127 4858: contig of 1732 bp in length
* 4859 4958: gap of unknown length
* 6113 6212: gap of unknown length
* 6113 8591: contig of 2379 bp in length
* 8592 12150: contig of 3459 bp in length
* 12151 12250: gap of unknown length
* 1251 14857: contig of 2607 bp in length
* 14858 14957: gap of unknown length
* 14958 17651: contig of 2694 bp in length
* 17652 21862: contig of 4111 bp in length
* 21863 25380: gap of unknown length
* 25381 25480: contig of 3418 bp in length
* 25481 29178: gap of unknown length
* 29179 29278: contig of 3698 bp in length
* 29279 33316: gap of unknown length
* 33317 33416: contig of 4038 bp in length
* 33417 38012: gap of unknown length
* 38013 38112: gap of unknown length
* 38113 43104: contig of 4992 bp in length
* 43105 43204: gap of unknown length
* 43205 47574: contig of 4370 bp in length
* 47575 47674: gap of unknown length
* 47675 52632: contig of 4958 bp in length
* 52633 52732: gap of unknown length
* 52733 57058: contig of 4326 bp in length
* 57059 57158: gap of unknown length
* 57159 62021: contig of 4863 bp in length
* 62022 62121: gap of unknown length
* 62122 67329: contig of 5208 bp in length
* 67330 67429: gap of unknown length
* 67430 73286: gap of unknown length
* 73287 73386: contig of 5857 bp in length
* 73387 80609: gap of unknown length
* 80610 80709: contig of 7223 bp in length
* 80710 87143: gap of unknown length
* 87144 87243: contig of 6434 bp in length
* 87244 96034: gap of unknown length
* 96035 96134: contig of 8791 bp in length
* 96135 104640: gap of unknown length
* 104641 104740: contig of 8506 bp in length
* 104741 113799: gap of unknown length
* 113800 113899: contig of 9059 bp in length
* 113900 125284: gap of unknown length
* 125285 125384: contig of 11385 bp in length
* 125385 137762: gap of unknown length
* 137763 137862: contig of 12378 bp in length
* 137863 152504: gap of unknown length
* 152505 152604: contig of 14642 bp in length
* 152605 168992: gap of unknown length
* 168993 169093: contig of 16388 bp in length
* 169093 194576: gap of unknown length
* 194576 194576: contig of 25484 bp in length.

```

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-763E3"
1..1065
misc_feature
/feature="assembly_name:Contig15"
1166..3026
misc_feature
/feature="assembly_name:Contig16"
3127..4858
misc_feature
/feature="assembly_name:Contig17"
4959..6112
misc_feature
/feature="assembly_name:Contig18"
6213..8591
misc_feature
/feature="assembly_name:Contig19"
8592..12150
misc_feature
/feature="assembly_name:Contig20"
12151..14857
misc_feature
/feature="assembly_name:Contig21"
14858..17651
misc_feature
/feature="assembly_name:Contig22"
17652..21862
misc_feature
/feature="assembly_name:Contig23"
21863..25380
misc_feature
/feature="assembly_name:Contig24"
25381..29178
misc_feature
/feature="assembly_name:Contig25"
29179..33316
misc_feature
/feature="assembly_name:Contig26"
33317..38012
misc_feature
/feature="assembly_name:Contig27"
38013..43104
misc_feature
/feature="assembly_name:Contig28"
43105..47574
misc_feature
/feature="assembly_name:Contig29"
47575..52632
misc_feature
/feature="assembly_name:Contig30"
52633..57058
misc_feature
/feature="assembly_name:Contig31"
57059..62021
misc_feature
/feature="assembly_name:Contig32"
62022..67329
misc_feature
/feature="assembly_name:Contig33"
67330..73286
misc_feature
/feature="assembly_name:Contig34"
73287..80609
misc_feature
/feature="assembly_name:Contig35"
80610..87143
misc_feature
/feature="assembly_name:Contig36"
87144..96034
misc_feature
/feature="assembly_name:Contig37"
96035..104640
misc_feature
/feature="assembly_name:Contig38"
104641..113799
misc_feature
/feature="assembly_name:Contig39"
113800..125284
misc_feature
/feature="assembly_name:Contig40"
125285..137762
misc_feature
/feature="assembly_name:Contig41"
137763..152504
misc_feature
/feature="assembly_name:Contig42"
152505..168992
misc_feature
/feature="assembly_name:Contig43"
168993..194576
misc_feature
/feature="assembly_name:Contig44"
194577..194576
misc_feature
/feature="assembly_name:Contig45"
194577..194576
misc_feature
/feature="assembly_name:Contig46"
194577..194576
misc_feature
/feature="assembly_name:Contig47"
194577..194576
misc_feature
/feature="assembly_name:Contig48"
194577..194576
misc_feature
/feature="assembly_name:Contig49"
194577..194576
misc_feature
/feature="assembly_name:Contig50"
194577..194576
misc_feature
/feature="assembly_name:Contig51"
194577..194576
misc_feature
/feature="assembly_name:Contig52"
194577..194576
misc_feature
/feature="assembly_name:Contig53"
194577..194576
misc_feature
/feature="assembly_name:Contig54"
194577..194576
misc_feature
/feature="assembly_name:Contig55"
194577..194576
misc_feature
/feature="assembly_name:Contig56"
194577..194576
misc_feature
/feature="assembly_name:Contig57"
194577..194576
misc_feature
/feature="assembly_name:Contig58"
194577..194576
misc_feature
/feature="assembly_name:Contig59"
194577..194576
misc_feature
/feature="assembly_name:Contig60"
194577..194576
misc_feature
/feature="assembly_name:Contig61"
194577..194576
misc_feature
/feature="assembly_name:Contig62"
194577..194576
misc_feature
/feature="assembly_name:Contig63"
194577..194576
misc_feature
/feature="assembly_name:Contig64"
194577..194576
misc_feature
/feature="assembly_name:Contig65"
194577..194576
misc_feature
/feature="assembly_name:Contig66"
194577..194576
misc_feature
/feature="assembly_name:Contig67"
194577..194576
misc_feature
/feature="assembly_name:Contig68"
194577..194576
misc_feature
/feature="assembly_name:Contig69"
194577..194576
misc_feature
/feature="assembly_name:Contig70"
194577..194576
misc_feature
/feature="assembly_name:Contig71"
194577..194576
misc_feature
/feature="assembly_name:Contig72"
194577..194576
misc_feature
/feature="assembly_name:Contig73"
194577..194576
misc_feature
/feature="assembly_name:Contig74"
194577..194576
misc_feature
/feature="assembly_name:Contig75"
194577..194576
misc_feature
/feature="assembly_name:Contig76"
194577..194576
misc_feature
/feature="assembly_name:Contig77"
194577..194576
misc_feature
/feature="assembly_name:Contig78"
194577..194576
misc_feature
/feature="assembly_name:Contig79"
194577..194576
misc_feature
/feature="assembly_name:Contig80"
194577..194576
misc_feature
/feature="assembly_name:Contig81"
194577..194576
misc_feature
/feature="assembly_name:Contig82"
194577..194576
misc_feature
/feature="assembly_name:Contig83"
194577..194576
misc_feature
/feature="assembly_name:Contig84"
194577..194576
misc_feature
/feature="assembly_name:Contig85"
194577..194576
misc_feature
/feature="assembly_name:Contig86"
194577..194576
misc_feature
/feature="assembly_name:Contig87"
194577..194576
misc_feature
/feature="assembly_name:Contig88"
194577..194576
misc_feature
/feature="assembly_name:Contig89"
194577..194576
misc_feature
/feature="assembly_name:Contig90"
194577..194576
misc_feature
/feature="assembly_name:Contig91"
194577..194576
misc_feature
/feature="assembly_name:Contig92"
194577..194576
misc_feature
/feature="assembly_name:Contig93"
194577..194576
misc_feature
/feature="assembly_name:Contig94"
194577..194576
misc_feature
/feature="assembly_name:Contig95"
194577..194576
misc_feature
/feature="assembly_name:Contig96"
194577..194576
misc_feature
/feature="assembly_name:Contig97"
194577..194576
misc_feature
/feature="assembly_name:Contig98"
194577..194576
misc_feature
/feature="assembly_name:Contig99"
194577..194576
misc_feature
/feature="assembly_name:Contig100"
194577..194576

```

```

misc_feature 8692..12150
/feature="assembly_name:Contig13"
misc_feature 12251..14857
/feature="assembly_name:Contig14"
misc_feature 14958..17651
/feature="assembly_name:Contig15"
misc_feature 17752..21862
/feature="assembly_name:Contig16"
misc_feature 21963..25380
/feature="assembly_name:Contig17"
misc_feature 25481..29178
/feature="assembly_name:Contig18"
misc_feature 29279..33316
/feature="assembly_name:Contig19"
misc_feature 33417..38012
/feature="assembly_name:Contig20"
misc_feature 38113..43104
/feature="assembly_name:Contig21"
misc_feature 43205..47574
/feature="assembly_name:Contig22"
misc_feature 47675..52632
/feature="assembly_name:Contig23"
misc_feature 52733..57058
/feature="assembly_name:Contig24"
misc_feature 57159..62021
/feature="assembly_name:Contig25"
misc_feature 62122..67329
/feature="assembly_name:Contig26"
misc_feature 67430..73286
/feature="assembly_name:Contig27"
misc_feature 73387..80609
/feature="assembly_name:Contig28"
misc_feature 80710..87143
/feature="assembly_name:Contig29"
misc_feature 87244..96034
/feature="assembly_name:Contig30"
misc_feature 96135..104640
/feature="assembly_name:Contig31"
misc_feature 104741..113799
/feature="assembly_name:Contig32"
misc_feature 113900..125284
/feature="assembly_name:Contig33"
misc_feature 125385..137762
/feature="assembly_name:Contig34"
misc_feature 137863..152504
/feature="assembly_name:Contig35"
misc_feature 152605..168992
/feature="assembly_name:Contig36"
misc_feature 169093..194576
/feature="assembly_name:Contig37"
misc_feature 194577..194576
/feature="assembly_name:Contig38"
BASE COUNT 60015 a 38425 c 37951 g 55254 t 2931 others
ORIGIN
Query Match 79.2% Score 19.8: DB 2: Length 194576;
Best Local Similarity 91.3% Pred: NO. 21;
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 2 acatgtcgcaagtcacgaatt 24
|||||
Db 155577 ACATTGTGCGCAATGACAGCAATT 155555

```

```

RESULT 24
AC012205/c AC012205 228434 bp DNA HTG 30-MAR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-340K9 map 4, WORKING DRAFT
DEFINITION SEQUENCE, 24 unordered pieces.
ACCESSION AC012205
VERSION AC012205.3 GI:7341726
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

RESULT	22
AC090618/c	LOCUS
DEFINITION	AC090618.4 GI:14718350
ACCESSION	AC090618
VERSION	AC090618.4
KEYWORDS	HTGS; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 193661)
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lande,E.
JOURNAL	Homo sapiens chromosome 17, clone RP11-763E3
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 193661)
Birren,B., Linton,L., Nusbaum,C., Lande,E., Allen,N., Anderson,S., Baine,N., Bastien,V., Boguslavsky,L., Bouknight,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArnell,R., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glend,S., Goette,M., Graham,L., Grand-Pierre,N., Hages,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrum,J., Menues,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Plunkhard,P., Pierre,N., Pollara,V., Raymond,C., Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severly,P., Sougez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J.J., Testaye,S., Theodore,J., Travers,N., Travis,N., Tribilio,J., Vassiliev,H., Vel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.	Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 1420 Charles Street, Cambridge, MA 02141, USA On Jul 14, 2001 this sequence version replaced gi:14547801. All repeats were identified using RepeatMasker: Smit,A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence.submissions@genome.wi.mit.edu Project Information Center project name: L12048 Center clone name: 763_E_3	
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. * This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved. * 1 25016: contig of 25016 bp in length * 25017 25116: gap of 100 bp * 25117 193661: contig of 168545 bp in length.	
FEATURES Source	1..193661

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-763E3"
/clone.lib="RPC1-11 Human Male BAC"
BASE COUNT      58274 a 38105 c 38669 g 58513 t      100 others
ORIGIN

Query Match      79.2%; Score 19.8; DB 2; Length 193661;
Best Local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      2      acattgtcgaagtcacgaatt 24
|||||
Db 120332 ACATTGTCGAATGACAGAAATT 120310

RESULT 23
AC019147/c
LOCUS      AC019147      194576 bp      DNA
DEFINITION      Homo sapiens chromosome 17 clone RP11-763E3, WORKING DRAFT
AC019147
SEQUENCE, 30 unordered pieces.
AC019147
AC019147.4 GI:8570396
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 194576)
Waterston,R.H.
The sequence of Homo sapiens clone
unpublished
2 (bases 1 to 194576)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 17, 2000 this sequence version replaced gi:7235343.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0763B03
----- Summary Statistics -----
Sequencing vector: M13; 85%
Sequencing vector: plasmid; 15%
Chemistry: Dye-primer ET; 85% of reads
Chemistry: Dye-terminator Big Dye; 15% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168961 bases at least Q40
Consensus quality: 176420 bases at least Q30
Consensus quality: 180507 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191676; sum-of-configs
Quality coverage: 3.44 in Q20 bases; agarose-fp
Quality coverage: 3.31 in Q20 bases; sum-of-configs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      1065: contig of 1065 bp in length
*      1066      1165: gap of unknown length

```

```
-----
Summary Statistics
Consensus quality: 135192 bases at least Q40
Consensus quality: 142393 bases at least Q30
Consensus quality: 143134 bases at least Q20
Estimated insert size: 180650; agarose-fp estimation
Estimated insert size: 144395; sum-of-contigs estimation
Quality coverage: 9.95 in Q20 bases; agarose-fp estimation
Quality coverage: 12.45 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1112: contig of 1112 bp in length
1113 1212: gap of unknown length
1213 2797: contig of 1585 bp in length
2798 2697: gap of unknown length
5106 5205: contig of 2208 bp in length
5206 8117: contig of 2912 bp in length
8118 8217: gap of unknown length
14132 14132: contig of 5915 bp in length
14133 14232: gap of unknown length
14233 19887: contig of 5655 bp in length
19888 19987: gap of unknown length
36187 36287: gap of 16200 bp in length
36288 51471: contig of 15184 bp in length
51472 51571: gap of unknown length
51572 67249: contig of 15678 bp in length
67250 67349: gap of unknown length
81383 81482: contig of 14033 bp in length
81483 104806: contig of 23324 bp in length
104807 104907: gap of unknown length
104907 145495: contig of 40589 bp in length.
Location/Qualifiers
1..145495
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone_lib="RP11-370B10"
/clone_lib="RP11-370B10"
BASE COUNT 45702 a 26543 c 26178 g 45970 t 1102 others
ORIGIN
Query Match 79.2%; Score 19.8; DB 2; Length 145495;
Best local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 acattgtcgcagtcacagaatt 24
||||||| ||||| ||||| |||||
Db 61367 ACATTGTCACAGTCACAGAATT 61345

RESULT 21
AC093207 186735 bp DNA HTG 16-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-448F14, WORKING DRAFT SEQUENCE,
13 unordered pieces.
AC093207
AC093207.1 GI:151933341
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 186735)
```

```
AUTHORS
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE Unpublished
2 (bases 1 to 186735)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 357346
Center clone name: CIT-HSPC_448F14
-----
Summary Statistics
Consensus quality: 174327 bases at least Q40
Consensus quality: 181558 bases at least Q30
Consensus quality: 182476 bases at least Q20
Estimated insert size: 109000; agarose-fp estimation
Estimated insert size: 185535; sum-of-contigs estimation
Quality coverage: 9.78 in Q20 bases; agarose-fp estimation
Quality coverage: 5.75 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1177: contig of 1177 bp in length
1178 1277: gap of unknown length
1278 2818: contig of 1541 bp in length
2819 2918: gap of unknown length
2919 5193: contig of 2275 bp in length
5194 5293: gap of unknown length
5294 8311: contig of 3018 bp in length
8312 8411: gap of unknown length
8411 11450: contig of 3039 bp in length
11451 11551: gap of unknown length
11551 16750: contig of 5200 bp in length
16751 16850: gap of unknown length
16851 27635: contig of 10785 bp in length
27636 27735: gap of unknown length
27736 40931: contig of 13196 bp in length
40932 41031: gap of unknown length
41032 55320: contig of 14289 bp in length
55321 55420: gap of unknown length
55421 76531: contig of 21111 bp in length
76532 76631: gap of unknown length
76632 108434: contig of 31803 bp in length
108435 108534: gap of unknown length
108535 135350: contig of 26816 bp in length
135351 135451: gap of unknown length
135451 186735: contig of 51285 bp in length.
Location/Qualifiers
1..186735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone_lib="CIT-448F14"
/clone_lib="CIT-448F14"
BASE COUNT 60409 a 32285 c 33947 g 58883 t 1211 others
ORIGIN
Query Match 79.2%; Score 19.8; DB 2; Length 186735;
Best local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```



```

repeat_region /rpl_family="MIR"
5716. .5913
repeat_region /rpl_family="MIR"
5983. .6010
repeat_region /rpl_family="AT_rich"
6039. .6272
repeat_region /rpl_family="L1"
6476. .7408
repeat_region /rpl_family="L1"
12965. .13128
repeat_region /rpl_family="L1"
13129. .13445
repeat_region /rpl_family="Alu"
13446. .13999
repeat_region /rpl_family="L1"
14000. .14386
repeat_region /rpl_family="L1"
14387. .14667
repeat_region /rpl_family="Alu"
14668. .14723
repeat_region /rpl_family="L1"
14724. .15471
repeat_region /rpl_family="L1"
15473. .15608
repeat_region /rpl_family="L1"
15639. .18141
repeat_region /rpl_family="L1"
18142. .18489
repeat_region /rpl_family="MalR"
18642. .18662
repeat_region /rpl_family="AT_rich"
20137. .20494
repeat_region /rpl_family="MalR"
20500. .20566
repeat_region /rpl_family="AT_rich"
21874. .22216
repeat_region /rpl_family="MERL_type"
22478. .22536
repeat_region /rpl_family="L2"
22712. .23154
repeat_region /rpl_family="ERV1"
23618. .23902
repeat_region /rpl_family="L1"
24035. .24055
repeat_region /rpl_family="AT_rich"
24243. .24487
repeat_region /rpl_family="MIR"
24807. .25180
repeat_region /rpl_family="MalR"
25519. .25816
repeat_region /rpl_family="Alu"
25904. .25966
repeat_region /rpl_family="AT_rich"
26064. .26379
repeat_region /rpl_family="Alu"
32986. .33157
repeat_region /rpl_family="ERV1"
33404. .33690
repeat_region /rpl_family="L1"
35170. .36476
repeat_region /rpl_family="L1"
37329. .37359
repeat_region /rpl_family="AT_rich"
37537. .37566
repeat_region /rpl_family="AT_rich"
37756. .38061
repeat_region /rpl_family="Alu"
38264. .38503
repeat_region /rpl_family="ERV1"
38545. .38585
repeat_region /rpl_family="L2"
39025. .39244
repeat_region /rpl_family="ERV1"

```

```

repeat_region 39413. .39589
/rpl_family="ERV1"
repeat_region 39782. .39946
/rpl_family="MIR"
repeat_region 40393. .40665
/rpl_family="Alu"
repeat_region 41151. .41855
/rpl_family="ERV1"
repeat_region 41945. .42232
/rpl_family="Alu"
repeat_region 42828. .43210
/rpl_family="ERV1"
repeat_region 43211. .43510
/rpl_family="L1"
repeat_region 43511. .43812
/rpl_family="ERV1"
repeat_region 44561. .44632
/rpl_family="L2"
repeat_region 44633. .45357
/rpl_family="ERV1"
repeat_region 45358. .46188
/rpl_family="L2"
repeat_region 46279. .46514
/rpl_family="Alu"
repeat_region 46874. .46958
/rpl_family="MalR"
repeat_region 47176. .48678
/rpl_family="L1"
repeat_region 53308. .54257
/rpl_family="L1"
repeat_region 54292. .54585
/rpl_family="ERV1"
repeat_region 54601. .54743
/rpl_family="L1"

```

Query Match 79.2%; Score 19.8; DB 9; Length 144355;
 Best Local Similarity 91.3%; Pred. No. 21;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatt 24

Db 107136 ACATTGTCTCACAGTGACAGATT 107158

```

RESULT 20
AC093273/C DNA HTG 16-AUG-2001
LOCUS Homo sapiens chromosome 5 clone RP11-370B10, WORKING DRAFT
DEFINITION
AC093273
AC093273.1 GI:15193407
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 145495)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 145495)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 548479
Center clone name: RPCI-11_370B10

```

AUTHORS Entcheva, P., Liebl, W. and Streif, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2001) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES Location/Qualifiers
 source
 1. 965
 /organism="uncultured bacterium pCosHE1"
 /db_xref="taxon:143796"
 /clone="pCosHE1"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(<1. .438)
 /gene="bioA"
 /gene="bioA"
 complement(<1. .438)
 /note="7.8-diaminopelargonic acid
 synthetase-aminotransferase; similar to bioA"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase"
 /protein_id="AAG60563.1"
 /db_xref="GI:12620105"
 /translation="MTQDDIAFDQDHIMPIYTSMTREPVYPVPAAXAHCKRRRLVD
 GSSSWMAATHCYHNPRLNAAMKAQTFQMSHVMEDGITRQAVADLCRRXVAMTPDALEC
 VPLDSGVAXEXXDENGASTAPLPADRPPEKYHAFPIPLIQ"
 525. .>965
 /gene="bioB"
 525. .>965
 /gene="bioB"
 /note="similar to bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AAG60564.1"
 /db_xref="GI:12620106"
 /translation="MXHHAARTMSQVTELEFNKPELLEMEAOOVHROHEDPRHYVOST
 LSLKIGACPEDCCKYCPQSAARYKGTGSESLMEVQVLDLSAKAKAAGSTRGMSGAAW
 KKHPIYTCYELQMWKGVKEMCLEALYDPRHARRAXXSACILIA"
 BASE COUNT 230 a 256 c 285 g 180 t 14 others
 ORIGIN
 Query Match 80.8%; Score 20.2; DB 1; Length 965;
 Best Local Similarity 88.0%; Pred. No. 13;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 gacattgtcgcaagtcacagaatta 25
 111 1111111111111111111111
 Db 545 GACGATGTCGCAAGTCACTGAATTA 569
 RESULT 19
 LOCUS AC016743 144355 bp DNA PRI 09-AUG-2001
 DEFINITION Homo sapiens BAC clone RP11-475H17 from 2, complete sequence.
 AC016743
 VERSION AC016743.10 GI:13443262
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 144355)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome research. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 REFERENCE 2 (bases 1 to 144355)
 AUTHORS Waligorski, J., Abbott, A. and Boyer, E.
 TITLE The sequence of Homo sapiens BAC clone RP11-475H17
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 144355)
 AUTHORS Waterston, R.H.

TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 144355)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 144355)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 ON Mar 24, 2001 this sequence version replaced gi:13270791.
 COMMENT
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0475H17

 NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.
 MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>
 SOURCE INFORMATION:
 The RP11-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
 Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute
 (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-471D6; the clone sequenced
 to the right is RP11-479L11, 200 bp overlap. Actual end of this
 clone is at base position 49393 of RP11-479L11.
 FEATURES
 source
 1. 144355
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-475H17"
 /clone_lib="RP11-11"
 403. .445
 /rpl_family="AT-rich"
 repeat_region
 1826. .2062
 /rpl_family="ERV1"
 repeat_region
 3399. .3600
 /rpl_family="MALR"
 repeat_region
 4181. .4244

QY 1 gacattgtcgcagtcacagatt 25
|||||
Db 4049 GACATTGTCCGACAGTCACATT 4073

RESULT 17
AC084797
LOCUS AC084797 175777 bp DNA HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-6203, WORKING DRAFT SEQUENCE,
28 unordered pieces.
AC084797
AC084797.2 GI:13786426
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 175777)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175777)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 25, 2001 this sequence version replaced gi:1192125.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
Project Information
Center Project Name: 0
Center clone name: RPCT-11_6203

Summary Statistics
Consensus quality: 131005 bases at least Q40
Consensus quality: 144992 bases at least Q30
Consensus quality: 151270 bases at least Q20
Estimated insert size: 158300; agarose-tp estimation
Quality coverage: 4.66 in Q20 bases; agarose-tp estimation
Quality coverage: 4.26 in Q20 bases; sum-of-configs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 configs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the configs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1140: contig of 1140 bp in length
1141 1240: gap of unknown length
1241 2356: contig of 1116 bp in length
2357 2456: gap of unknown length
2457 3574: contig of 1118 bp in length
3575 3674: gap of unknown length
3675 4725: contig of 1051 bp in length
4726 4825: gap of unknown length
4825 6209: contig of 1384 bp in length
6210 6309: gap of unknown length
6310 7534: contig of 1225 bp in length
7535 7634: gap of unknown length
7635 8906: contig of 1272 bp in length
8907 9006: gap of unknown length
9007 10031: contig of 1025 bp in length
10032 10131: gap of unknown length
10132 11204: contig of 1073 bp in length
11205 11304: gap of unknown length
11305 12612: contig of 1308 bp in length
12613 12712: gap of unknown length
12713 13858: contig of 1146 bp in length

* 13859 13958: gap of unknown length
* 13959 17120: contig of 3162 bp in length
* 17121 17220: gap of unknown length
* 17221 19014: contig of 1794 bp in length
* 19015 19114: gap of unknown length
* 19115 23039: contig of 3925 bp in length
* 23040 23139: gap of unknown length
* 23140 25203: contig of 2064 bp in length
* 25204 25303: gap of unknown length
* 25304 28557: contig of 3254 bp in length
* 28558 28657: gap of unknown length
* 28658 32217: contig of 3560 bp in length
* 32218 32317: gap of unknown length
* 32318 39892: contig of 7575 bp in length
* 39892 39992: gap of unknown length
* 39993 47482: contig of 7490 bp in length
* 47483 47582: gap of unknown length
* 47583 55851: contig of 8269 bp in length
* 55852 55952: gap of unknown length
* 55953 62694: contig of 6742 bp in length
* 62694 62794: gap of unknown length
* 62794 74490: contig of 11697 bp in length
* 74491 74590: gap of unknown length
* 74591 82770: contig of 8180 bp in length
* 82771 82870: gap of unknown length
* 82871 95653: contig of 12783 bp in length
* 95654 95753: gap of unknown length
* 95754 108415: contig of 12662 bp in length
* 108416 108515: gap of unknown length
* 108516 126483: contig of 17968 bp in length
* 126484 126583: gap of unknown length
* 126584 145426: contig of 18843 bp in length
* 145427 145526: gap of unknown length
* 145527 175777: contig of 30251 bp in length.

FEATURES
source
1. 175777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-6203"
/clone_11b="RPCT human BAC library 11"
BASE COUNT 49834 a 38187 c 38793 g 44806 t 4157 others
ORIGIN

Query Match 85.6%; Score 21.4; DB 2; Length 175777;
Best Local Similarity 95.7%; Pred. No. 3.1;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 acattgtcgcagtcacagatt 24
|||||
Db 51444 ACATTGTCCGACAGTCACAGATT 51466

RESULT 18
AF250770
LOCUS AF250770 965 bp DNA BCT 31-JAN-2001
DEFINITION Uncultured bacterium PCOSHEL DARA-aminotransferase (btoa) and
biotin synthase (btoa) genes, partial cds.
ACCESSION AF250770
VERSION AF250770.1 GI:12620104
KEYWORDS
SOURCE
ORGANISM
uncultured bacterium PCOSHEL.
Bacteria; enviromental samples.
REFERENCE 1 (bases 1 to 965)
AUTHORS Entcheva, P., Liebel, W., Johann, A., Hartsch, T. and Streif, W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196
PUBMED 1133432
REFERENCE 2 (bases 1 to 965)

```

CDS
    /gene="hutu"
    <1..502
    /gene="hutu"
    /note="putative urocanate hydratase"
    /codon_start=2
    /transl_table=1
    /product="urocanase-like protein"
    /protein_id="AAG53593.1"
    /db_xref="GI:12407618"
    /translation="PRETISFQGLPARICWGLEWROKLGAFENWVRREGEVAPTIV
GRDHDSGVSASPNEFEMOGSDAVSPMLLNLNTASGATWSLHHGGVCGMF
SOHAGVIVCDGTDAAARLARVLRHNDPATGVRRHADAGYDIAIDCAKEQGLNLPRIA
ATOGOH"
    502..2031
    /gene="hutu"
    502..2031
    /gene="hutu"
    /note="putative"
    /codon_start=1
    /transl_table=11
    /product="histidine ammonia-lyase-like protein"
    /protein_id="AAG53586.1"
    /db_xref="GI:12407611"
    /translation="MAATITPTGCTITLAQLROYWQOPLQITLDESAHEAIVNACVE
AIYABGRTAGINTGRTGGLAOTRIATHDLENQRSIVLSHAACVCEPDDDDIVRLMNV
LKTNSLRGFSGLRSLVIOALIALVNAVGYSDPPASVSGASODLAPLHMSLFTLGE
GKARYRGEMLPATLALQKALAPVTLAEKGLALNGTQASTAFALRGLEEDLFAS
AVYRGALITTEAVIGSRPPDARLHEVRGQGLIDAALEFHVLTDTSAIEASHNDK
VDDPSLRQOPQVAGCLTQMRQVAEVLVLESNAVDNPLVFAENEMEVFACVTAAL
VAMADNLALAIETGALSRRIRALMMQDKHNSQLPPLVRNGSGNEMFACVTAAL
ASNKGLCHPTTSVDPKPCRSANQEDIVSMAPARRLMEAGNTRVLAENMLAAC
QCADLRGULTSSPLLEQARSGEYVAHTDDDFEAPDTEATLSLTKSGLVGLPAF
L"
    complement(2130..2606)
    /note="ORF1; similar to hypothetical 17.1 kDa protein in
modc-bioa intergenic region"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAG53587.1"
    /db_xref="GI:12407612"
    /translation="MKLISODLRDGLKPLRHVFNMGYEGDNISPHLAMDVPSTGK
SEVVCYDPAPDPSGMMHVIYANIPADTRVLPQSGSSLVALPEGAVTRIDPKKAG
YGGAPPKGETHXYITFVAHINVERIEVEEASGAVGVNHPHTLGSASITAKMS"
    complement(2651..3940)
    /gene="bioa"
    /complement(2651..3940)
    /note="7,8-diamino-pelargonic acid aminotransferase"
    /codon_start=1
    /transl_table=11
    /product="DAPA aminotransferase Bioa"
    /protein_id="AAG53588.1"
    /db_xref="GI:12407613"
    /translation="MTTDLAFDRQHTMHPYTSMTQPLFVYPVVEAHGCELTLAGSEQ
LYDGMSSWMAAIIHGYNHPRIANAMKAQIDRMISHVFGIITHQPAVDICRLVAMPET
LECVRPADSGSAVEYAKMALQYHAKGESRORLTPRNGHCHDTFGESGSCDONS
MSLKKGYLPENLFAPAPQSRFDGAWDEDMVCIAPLMAHHEHTAAVLTLEIYOGAG
GMATYHPENLKRIKRCDEKGLILLADIETATGRTGKLPACBIHAGITPDIICLGKAL
TGGTHTMSATLTTROVAETTSNGEACGFHNGTTPMGNPLACVASESTALLFSGEMOD
OVAATFESQLOEELAVARDEFPVADVRLGAGVETTHVNMMAALORFPVACVWVRP
FGKLIIMPVYVRPDLKSLTRAVADVAAVAHAHFR"
    4029..5069
    /gene="bioa"
    4029..5069
    /gene="bioa"
    /codon_start=1
    /transl_table=11
    /product="biotin synthase BioB"
    /protein_id="AAG53589.1"
    /db_xref="GI:12407614"
    /translation="MAHQARWTLISOYTALFEKRLLELPEAQOIIHROHFDPOOIOVST
LSTIKTGACPEDECKYCPQSARYKTGLESERLMEVEGVLESARQAKNAGSTRCKWCAAW
```

```

KNPERDMPVLEQWVQYKANGIQACMTLGLTDDHQORLASAGLDYYNHNLDPFPE
YGNLITTRTYOERDITLDRKADACITKCSGSIYGLCEFTVDRAGILLQNLNLPPEP
VPIINLVKVCSTPLANDNDVYAFDXIRIXSVAXYMXPTLVRLSTGREGONBOQAAC
FMAGNASTFVYCGKLLTTPNPEDEKDVQLFRRLGINTPQQTAVLTSDNQQHQLEQULFN
ADTDQFYNAATV"
5066..6223
/gene="bioc"
5066..6223
/gene="bioc"
/note="7-keto-8-amino-pelargonic acid synthetase; similar
to BioB"
/codon_start=1
/transl_table=11
/product="7-KAPA synthetase"
/protein_id="AAG53590.1"
/db_xref="GI:12407615"
/translation="MTMOORTDALDARRADALRVTPYENGAGRLIFRORRCMF
SGNDYIGLSQHPETIRMOGCAROYGVSGGSGVSGVTAHQAHLERHLAGMCGYPRAL
LFTISGRANQALIALTDKEDRIIVABRLSHASLLEKASHSPADLRFPARHNDVQDLS
LLDKPCHQOQLVTEGIFSMGDSAPRLGVTHAAQROGSMWLLVDHAGTGVTPGEGK
SAAQSGYKPELVVTEFGKGVSGAAVYCDPVDYILOFARHLIYSTSMPPQAQVAL
LAEFNVIRSDENARRORLALIDQFAGVADLPVQITASQSAIOPIVSENNRALHL
AQRRLREGIWVTAIRPPTVPAGTARLRLTILAHQPADITAIYALHVAQE"
6207..6962
/gene="bioc"
6207..6962
/gene="bioc"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis Bioc-like protein"
/protein_id="AAG53591.1"
/db_xref="GI:12407616"
/translation="MSOVNKOAIAAFGRAAOSYSRRDALOROSADGLRALLAENFPA
SYLDAGCGPGSNSLAWRQTSVYVAMDLCAPMLEEARNOAASUYLOADIEALPLADA
OESLASHLAVOXCASLPRALAEFLRVTOGKVAFTLLNGSLPELNOAMQAVDPP
HANRPLSGOQVNDALSGMOWYRCAYQVPTLLFDDALSMSHLKIGATHLHAGRAKRL
TRGLOQLAQAAMPQOACKFPLSYOLFHLIIRD"
6940..7680
/gene="biob"
6940..7680
/gene="biob"
/codon_start=1
/transl_table=11
/product="gethioiotin synthase Biob"
/protein_id="AAG53592.1"
/db_xref="GI:12407617"
/translation="MELSNVTEREYVTGTDEKGTAVASTALLQARALGRHATGYKP
VASSGEMTAQGLRNSDALADQRNSLALSYPAVPYFAEPTSPHIVASDEORPIEFA
TLSAGLRALFEOQAMVIVEGAGWFTPLSOPQFADYVMAERLPIVLVGVKGCINH
AMLIAQALRQAGLHFGAGVAVGVVPPKRAHEVWATLRVLSAPLGLGELPWLGBQADT
SAVCOYDIDTLALTPSAPSSGOSAGHPAR"
complement(7619..8227)
/gene="eisa"
complement(7619..8227)
/note="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-like protein"
/protein_id="AAG53594.1"
/db_xref="GI:12407619"
/translation="DCXLRTRXGENXIVSCITAGLLPIASGTWVWHEAPGAVNOAGTEI
XVPQDRITFSQLTVEENHIAQMAKGIAGSGISROVFEFFPELYTPKOLRAGLND
QOYOLALANALVTRPRLLIIDPSSRGQGLFTYKLIQOLLIRPNEELGLSVILAEQHL
LIRRYADRFCLLYGRNVAQGRVSELDPILAHMGFTASAO"
1512 a 2642 c 2605 g 1453 t 15 others
BASE COUNT
ORIGIN
```

Query Match 87.2% Score 21.8; DB 1; Length 8227;
Best Local Similarity 92.0%; Pred. No. 2;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

source
1. .5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(52. .528)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modC-bioA
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDCKLPHRHVFNQGYDGNISPHLAWDPVACTK
SFVVTYDPAIPGSGMMHVVNLPADTVLPGQSGLVAMPDGLVOTRDFGKAG
YDGAAPPKGETHRYIFTVALDIERIDVDEGASGAWGPNVPHSLASITAMS"
complement(587. .1876)
/gene="bioA"
complement(587. .1876)
/gene="bioA"
complement(587. .1876)
/note="7, 8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDLAFDORHIMHPYTSMTSPLYPVSAEGCELLSDGR
LVGMSWMAAIGHYNNPQILNAAKSOIAMSVMHGJTHAPALELCKRLVAMPPO
LKCVLADSGSVAEAMKALQYMOAKSGARQFLTFPNNGYIGDFGMSVCDPNS
MHSILMGYLPENLEFAPAPOSRMDGEMDERWVFALMAHREIAVILIEPVGAG
GMAYHPEWILKTRIKICDREGILLIDELATGSGRGKLPACFHAFLADICLCKAL
TGGTMTLSATLITRREVAERTISNGEACGPNHGPPFNKPLACAAANSLALISGDMPO
QVADIEVQLREQLAPADAEKADVAVLGLGVETTHFVMAALCKFEVEQSWIRP
FGRLYLMPPYIILIPOLQRLTAAVNAVODETFEQ"
1963. .3003
/gene="bioB"
1963. .3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPRWTLISQVTELFEPULDLLEAQQVHRHQHPPOVYST
LSTKTGACPEDECKYCPQSSRYKTGLAEERLMEVQLESARAKAAGSTRFCMGAM
KNPERMDPYLEOVGVKDLGLACWTLGILSESQYQRLANGLDLYINILDTSPF
YGNITTRTYOERLDTLEKVDAGIKVCSGIGLGEYDRAGLLLOLANLPTPEP
VPIMLVKGVTPLADNDVDADFETITIAVARIMPTSYVLSAGREOMNEQTOAMC
FMAGANSIFYGCKLITPPNEDEKDLQGLFKLGLINQQTAVLAGDNEQQLRLEQALMT
PDTDEYNAAL"
3000. .4154
/gene="bioC"
3000. .4154
/gene="bioC"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioC"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWQEKINAAIDARRAADALRRRYPVAGCGRWLVDDROYLNF
SSNDYGLSHHPQIRAMOGCARBOFGSGSGSHVSGSYVHQALEFEELAEWLGYSRA
LLPISGFAANDVIAAMAKEDRIADRSLSHASILEASISFSDLRFRFANDVTHLAR
LLASPCGQDLVTEGVFSMDGSAPIAEIQVYTOQHNGLMWDADAGTGVIGEQGS
SCHLQVAPPELLVTEFGKGVSGAAVLCSTVADYLLQFARHLISTSPAPQAOAL
RASLAVIRDEGDAERKELVSLAHFRAGYQDLPFTLADSCAIOPLIVGDNRSRALQ
AEKLRQGCWVTAIRPTVPAGTARLRLTLTAHEMDIDRLLEVLHGNC"
4141. .4896
/gene="bioc"
4141. .4896
/gene="bioc"
/gene="bioc"

```

```

/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein Bioc"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVKNQATIAAFGRAAALYEQHADLQROSADVLLAMLPERRY
HLDAGCGPMWRHREBRHNOVYALDSPMLVOARKDADHYLAGDIESPLATA
TFDLAMSLAYQWGCNLSIALRELRYVRERKGVAFPTLVQSGIPELHQAQVADRP
HANRLPPEDETESLNGVYHIOHLPITLTFDALSMSRLKIGATHHLEGDRLL
TRSQLQRLQIAMPQOQGRYPLTYHLGLVARE"
4883. .>5526
/gene="bioc"
4883. .>5526
/gene="bioc"
/note="DTB synthetase"
/codon_start=1
/transl_table=11
/product="gethiobiotin synthetase BiocD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYEYCTDTEWCKTVASCALLQAAKAGRTAGYKPVAS
GSEKTPBELRNSDALQORNSLQIDYATVAPVTFAPETSPHILISAOEGRPISSVMS
SGRLALQEQADWLVLEGAGWFTPLSDFTFADWVTOQLPVLLVGVKLGCLNHAWL
TAAQIQHAGTLVAGVANDVTPGKRHAETITLRLMIPAPLGEIEMLAENDE"
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN

```

```

Query Match 93.6%; Score 23.4; DB 1; Length 5526;
Best Local Similarity 96.0%; Pred. No. 0.29;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacatgtgcgaagtgcagaatta 25
Db 1983 GACCTGTGCAGACGACAGAAITTA 2007

```

```

RESULT 16
AF248314
LOCUS
DEFINITION
AF248314 8227 bp DNA BCT 24-JAN-2001
uncultured bacterium pCosAS1 urocanase-like protein (hutu) gene,
partial cds; histidine ammonia-lyase-like protein (hutu), DAPA
aminotransferase BioA (bioc), biotin synthase BioB (biob), 7-KAPA
synthetase (bioc), biotin biosynthesis Bioc-like protein (bioc),
and dehydrobiotin synthase BioD (biob) genes; complete cds; ABC
transporter-like protein (elsa) gene; partial cds; and unknown
gene.
ACCESSION
AF248314
AF248314.1 GI:12407610
VERSION
1 (bases 1 to 8227)
SOURCE
uncultured bacterium pCosAS1.
ORGANISM
Bacteria: environmental samples.
REFERENCE
1 (bases 1 to 8227)
Batscheva, P., Liebl, W., Johann, A., Hartsch, T. and Strelt, W.R.
Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL
11133432
PUBMED
2 (bases 1 to 8227)
Batscheva, P., Liebl, W. and Strelt, W.R.
Direct Submissio
Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
source
1. .8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502

```

gene

/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNKRVICCGFTSGCPHSPSOAIAVCGIKMVGATPRK;
CTHLCIPVNTREAVNAATGATASVTVYAPRCKDSILKIDAGIKLITTEGIFT
LMLITTVKVLDEAGVMTGPRCPGVITTPGCKKIGIQPHHKPGKVAIVRSSTLYE
AVKOTTYGFGQSTCVGIGSDPIPGSNFIDILMFEPKDPQTEATVMIGEGSSAEBA
AAVTKHEVTKPVVGVYAGVTAPKGRMGHAGAIAGGKGTADKFALEAGVKTVA
LADIGALKTKVLK"
2781..3686
/gene="ECs0755"
2781..3686
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
Escherichia coli g1417043|sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELITVFEVVAHRESYTRAFLALTQSAVFRQVSL
BEIRTPFESHKSKRIPLNDTGKYYLGIVKETLNKLEBDNTIINTWQPVQVLEAVN
PTFSTWMLIPNLHEPKLHPDITVINHSLANNDDINREYDAVIMKENCAPAEVEY
LFEELIPVCSGLAMSDQKLSVAELITELPLHOSRTITGWEENFALSGVSPVYN
NGPRFPLISMLIAAVSNLGLALLPRFALIDHDSDMVIPOCVPIRTGGRFIMTQCE
EKSDPHLOOFREMLAKSVVPOEM"
complement(3720..4322)
/gene="ECs0756"
complement(3720..4322)
/note="probable cob(1)amin adenosyltransferase, similar
to cob(1)amin adenosyltransferases (cortinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P30401|BTRF_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)amin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTERYVQRQOKLAEQVPTRVAAATEKKGILVPTGNGK
KSTVAFGTVIRAVHGKTVGVAOYIKGMDNGEYNLQPLGVEFHMGTFWETQNR
QADIDAEKVEVSESKRMADKRYDLVULDELTYMLAYHYLDTEEVIASLQNNPQQSV
IVRGCHSOILKMDATVEIRPVKHAHPNGIQVQSGIOW"
complement(4332..5984)
/gene="ECs0757"
complement(4332..5984)
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
[Escherichia coli
g11205981|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPTWQEPFLQNKDGTVEYLISDQHITVTELDGEVYKTAPE
ALVLLARQAFEASEFLKPSRHLQOVAASILLNDQASNNKYVALQILRAAEVSAKVL
NODDTGATIVASKGQIWTJCNDAEALSKGIYTFQENNLRFSGNAPLMDYEVNTQ
TNLPAQIDISAVAGDEYHLJYCNKGGGSAKALYQETKSLQPEKLTAFILIEKKSL
GTAACPPIHIAFVVGELSDQFLTKAKLASTKYNNLPSGMEGQCARDEILEVLL
EASQPGICAGRGCKYFAHDIVITLPRIGSGCPTAMALSCADNITAKIKKHGIWL
KLEHNPGYIFASLRENNHQAQHVQULDRPLRDVQDLARLPLVGTAVSLSCPTIVAR

DIHAKKIKARLDGSEPMPEYLKHHIVVYAGPAKTPENNACSLAGTTGGRMDGYIDFF
QAAAGSLVMSKGNRSQGVTDACHKHGFGNIGSGAALLAQEVKSLRCLLEYBELG
MEAVWMEYENLPAFLIVDDKGNFNSQFEDQHRKASCAPAH"
complement(6092..7372)
/gene="ECs0758"
complement(6092..7372)
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671|sp|P24943|GLUT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKISLTVMILALVIGMIIGVNLNTASPEAKLYAQEISIFP
TIELRLKMIILAVLVSITLVVGIAMKGDAMKAGIFSKTFLFICASLISIALGLITV
NFMFGTGIFNEVAGHAEFTGVAAEPTLKVFI SHAPESTIVDAMAHNEILQIVFSL
FLGCSITFAIGEGSAIVHALDSLAAHMLKLTGYVLPAPLTPAASALAEGLAVM
VSAIGEMGEFFETVLLWLLIGLIVVGCPIRRLPAISEPALAFPTSSSRAAP
GTLKELPOGVSPKIASPVLPIGVSPNVGSMATCSPATVITACACNHLSIGQITM
LILMLTSGKAGVPRASRVVIAATLNOFNIPEGLIILMGVDFLDMGRSATVWMSN
AMGAAMVSRWSEHFGEGCRGKALKPNESVALP"
complement(7533..7850)
/gene="ECs0759"
complement(7533..7850)
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 100.0% Score 25; DB 1; Length 297816;
Best Local Similarity 100.0% Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 gacatgtcgcaagtcacagaatta 25
|||||
Db 92331 GACATTGTGCCAAGTCACAGAAATTA 92355
RESULT 15
AF250776
LOCUS
DEFINITION
Oy 1 gacatgtcgcaagtcacagaatta 25
|||||
Db 92331 GACATTGTGCCAAGTCACAGAAATTA 92355
AF250776 5526 bp DNA BCT 31-JAN-2001
uncultured bacterium pcosHE2
Bacteria: environmental samples.
1 (bases 1 to 5526)
Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W.R.
isolation of complete operons and genes from microbial consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
2 (bases 1 to 5526)
Entcheva, P., Liebl, W. and Streif, W.R.
Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
Location/Qualifiers

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Polamoussis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R.

TITLE Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

JOURNAL Nature 409 (6819), 529-533 (2001)

MEDLINE 21074935

PUBMED 11206551

REFERENCE 2 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Polamoussis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source location/Qualifiers

1..13501

/organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="enterohemorrhagic"

<1..7576

/note="O-island #36; Region of the EDL933 chromosome not homologous to *E. coli* K-12 M61655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"

66..665

/gene="lomK"

/note="20981"

66..665

/gene="lomK"

/function="putative membrane; Other or unknown (Phage or Prophage Related)"

/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from GenPept 118 : g117532789[gb]AA63231.1[AF151091]_2 (AF151091) lom [prophage P-ElbA]"

/codon_start=1

/transl_table=11

/product="putative outer membrane protein of prophage CP-933K"

/protein_id="AA655137.1"

/db_xref="GI:12513752"

/translation="MRKVCATIIISAAICLAVSGVPAMASEHSTLSAGYLAHSTADPAGSDLNGINVKYREFTDPLGLITFSYANAEDDOKTHYSDTRHEDYVNRHFSYVAGPSYRVNMFSAVMAGVAYSRYSTSGDYFRVDNKRKTHDVLGTSGSDARYSNTSLAWGAGVGNPTESAVVDVAYEXSGSDWRDGTGIVGYGR"

719..2041

/gene="20982"

719..2041

/gene="20982"

/function="putative structure; Structural component (Phage or Prophage Related)"

/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from GenPept 118 : g114585436[gb]AAD25464.1[AF125520]_59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"

/codon_start=1

/transl_table=11

/product="putative tail component of prophage CP-933K"

/protein_id="AA655138.1"

/db_xref="GI:12513753"

/translation="MGMVAVVISGVLLKDGAKPIQNCITQLKARNSTXYVVNTVASENFDEGRYSMDVEGYSVTLLEGFPSPHAGTISYEDSOPGTINDFLGAATFEDARPEALRFEQWVEEARHAEAKKNAGAEETSRNAGISASKASASANDTSASESSSAROAEEASAKKSEASSSSASFEACKASISLOSATDPAELSKTAASAGNARDAITTEKAREESASQASBOSRIAEADAVARIPYVGPPEKGPAGIQGKGGKG"

gene

CDS

2489..3469

/gene="20985"

2489..3469

/gene="20985"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from GenPept 118 : g1169603671[gb]AA633527.11 (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium 172j]"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655140.1"

/db_xref="GI:12513755"

/translation="MLSPIRTTFHNSVNIYVSSPCQTVSPAKYEELKVIDEKTPILEOMFEPNPERYKKEDEPIYNTKQHPYLDNVAIAEEDRMIGFVDFGDFVNTKATSKLEDFENVMYIYREDYDFESMDKLSLDIYHDTICBRLTEPKRDEYLIINLEKELR EISKAQDSLIMYAKKRNHNFDFRMLALKAETIRCTYNNKNGISGEGCITLDMDNLTGRLGTVAPDGI SMHVDNRNDSVNIENSALIVNSNHPALLEGSEFMHNSKVD AHPYVDIGKGVKRYFNFTPLAHNHCDFIEFNHPIIMNTSOYTCSSW"

3530..4522

/gene="20986"

3530..4522

/gene="20986"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655141.1"

/db_xref="GI:12513756"

/translation="MKTPSLQSNFNSAPAGYSAPAPNRAENAVDYVDIGKRIPLSAADLSNVYESVIRAVHDSRSLDQFTVMIGVAVDALSRQSTFADAVSYIINHEK VHCICIRYNEELNNESSVKIIDISLTNCELYVGOEPIFPCIEAGENENDEY VSFVAPDIDSYEMSWOEGIHETIIHVGSSPSDSDSITELGPEILARRAOEGL MSVPDFKCYAPEPEAHRLRLNLALQQAAMRHEENRAFERKLGITSDRYEASPDPT EYSAVNIIGTFEIQHDFPGIADINDNLDANOITOLYHGAYITFFGVDRHNOO"

5350..6231

/gene="20989"

5350..6231

/gene="20989"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655142.1"

/db_xref="GI:12513757"

/translation="MLSPYSVNLGCSWNSLITRNLITSPDNKRVLSVVDAAVHSDNGAOV

CDS
 complement(147. .1430)
 /gene="ybhC"
 /function="orf: Not classified"
 /note="fa27: 98 pct identical to fragment YBH_CECOLI
 SW:P46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative peclnesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /db_xref="GI:1786989"
 /translation="NMFSVSRALALAFGLTACSSPPDQPSDQAPCTSSKPI
 LSAEONFDAQHVFASLTPGAAMPSPITLPAQDFVGPAGTQGVHTTQAVD
 AAIKRTNRQYIAVMGEYGVYPAAGGITLXGCEKPIDVKIGLSLDCGSPA
 DMHEDYPRGKMPKPAWMDSCOSKRSIDGICSAFWSNNGLOLNTENT
 LGSVDAGNHPAVALRTDQDOINNVNLTGRNTPFVNSGVONRLEFTRPILVT
 NSTIEDVDIVSGRAVFPDNTFERNVNSTOQEAIVFPATLSNITTYFPLVNSFN
 AFQGVAAQGRSLVDANTNGQVIRDSAINEGFNAPKPADAVISNRPAGNCSVD
 DNDEIQRLNDITVYNNMMEYNMRGVSKEVAAKK"
 complement(1507. .1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 complement(1582. .2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582. .2058)
 /function="orf: unknown"
 /note="f158: 99 pct identical to YBH_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLDDGKLEHRRHVGKMGYDGNISPHLAMDVPACT
 SEVVTCTDPAPTGSGMHWVNVNLPADRVLPQGGSLVAMPDGLQTRIDFGTG
 YDGAAPKGETHRYIFTVALDIERIDVDEGASGAWGVNHFHSLASITAMS"
 complement(2108. .2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117. .3406)
 /gene="b10A"
 /note="b0774"
 complement(2117. .3406)
 /gene="b10A"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="fa29: 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDLAFDORHIMHPTSMSPILPYVSAEGCELLISDGR
 LECVSSWMAAIGHVYNNHNPOLIAAMKSOIDAMSHVGGITHAFAIILCKRLVAMP
 LECVFLADSGVAVVAMKALIOYNOAGKARPLFRNGYAGDFFGAMSCDDPNS
 MHSIWMGYLPENLFAPOASRMDEMDERMGVAFMAAHRHETAAVILPEVQAG
 GMRMYPEMLKRTKICDREGILLINDELTGTGPRGKLFACNATIAADILCLGAL
 TGGTMTLSATLITREVAETISNGEAGCFMGPFGKGNPLACAANASLALIESGDMQ
 QVADIEVQLREQLAPARDAEMVADVILGIVETHTHPVMAALQKFEVQGVWTRP
 FGRILYLMPPYILIPOLQKRTAAVNVMAVODEFFCQ"
 complement(2193. .2221)
 /gene="b10A"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411. .3450
 /bound_moiety="B10B documented site"
 3411. .3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_moiety="B10B predicted site"
 complement(3411. .3450)
 /note="central position to predicted promoter:50"
 protein_bind
 protein_bind
 protein_bind

protein_bind
 /bound_moiety="B10B predicted site"
 complement(3411. .3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="B10B documented site"
 3413. .3441
 /note="factor Sigma70; promoter bioB; documented +1 at
 808525"
 complement(3447. .3473)
 /note="factor Sigma70; promoter bioA; documented +1 at
 808515"
 3493. .4533
 /gene="b10B"
 /note="b0775"
 3493. .4533
 /gene="b10B"
 /EC_number="2.8.1.-"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o346: 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="Biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPMTLSQVTELEPERLLDLFEAQVYHROHPDPROYST
 LLSITGACPEDECKYCPSSRYKTGLAEERIMEYOYLESAKAKAGSTRFCGAA
 KNHEROMPYLEOMVQYKAMGLFACMTGLTSSQORLANAGLDVYNNHIDPSPE
 YGNTITRTYERLDITLEKVDACITKCSGGITVGLGTVRDAGLLIOLNIPRPS
 VPIMLVVKGSTPLADNDVDVDFPITRIARVIMPLSTYVRSACGNEBQTOAT
 FMAGNSLFYQCKLLTPNPEDEKDLQIFRLGLNPOOTAVLADNBQORLQALMT
 PDTDYVNAAL"
 4530. .5684
 /gene="b10F"
 /note="b0776"
 4530. .5684
 /gene="b10F"
 /EC_number="2.3.1.47"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o384: 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSQEKINALDARRADALRRYPPVACGGRVLVADDDRYLNF
 SSNDYLSHHPQIIRAWQGADEQFGTSGSGSVSGVYHQALEELAEWIGYSRA
 LLTISGFANQAVIAAMAKEDRIAADRLSHASLSLEAASLSPQLRRFANHDVTHLAR
 LIASPCQOMVTEGVFSGMDGDSAPLAELIOQVQOHNGWLMVDAGTGTGQGGRG
 SCWLQKVPPELLIVTFKGFVSGAVALCSSTVADYLLQFARHLITSTMPAQAAL

Query Match 100.0% Score 25; DB 1; Length 1102;
 Best local similarity 100.0%; Pred. No. 0.043;
 Matches 25; Conservative 0; Mismatches 0; Gaps 0;
 Oy 1 gacattgcgaagtcacagaatta 25
 ||||||||||||||||||||
 Db 3513 GACATTGTCGCAAGTCACAGAATTA 3537
 RESULT 13
 AE005258
 LOCUS
 DEFINITION
 OF 155.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 1 (bases 1 to 13501)
 AE005258 13501 bp DNA BCT 21-MAR-2001
 Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 OF 155.
 AE005258 AE005174
 AE005258.1 GI:12513751
 Escherichia coli O157:H7 EDL933.
 Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaactcacagaatta 25
    |||||||
Db 137 GACATTGTCGCAAGTCACAGCAATTA 161

RESULT 10
ARI01809 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 1 from patent US 6083712.
ACCESSION ARI01809
VERSION ARI01809.1 GI:12812607
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
    source
        1..5872
        /organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DH 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaactcacagaatta 25
    |||||||
Db 137 GACATTGTCGCAAGTCACAGCAATTA 161

RESULT 11
ARI01810 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 6 from patent US 6083712.
ACCESSION ARI01810
VERSION ARI01810.1 GI:12812608
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
    source
        1..5872
        /organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DH 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaactcacagaatta 25
    |||||||
Db 137 GACATTGTCGCAAGTCACAGCAATTA 161

RESULT 12
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
DEFINITION genome.
ACCESSION AE000180 U00096

```

```

VERSION AE000180.1 GI:1786988
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
        Escherichia.
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
        Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
        Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
        Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
        University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
        Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
        608-263-7459
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
        University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
        Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
        608-263-7459
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
        Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
        This sequence was determined by the E. coli Genome Project at the
        University of Wisconsin-Madison (Frederick R. Blattner, director).
        Supported by NIH grants HG00301 and HG01428 (from the Human Genome
        Project and NCHGR). The entire sequence was independently
        determined from E. coli K12 strain MG1655. Predicted open reading
        frames were determined using Genemark software, kindly supplied by
        Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
        30332 (e-mail: mark@amber.gatech.edu). Open reading frames that
        have been correlated with genetic loci are being annotated with CG
        Site Nos., unique ID nos. for the genes in the E. coli Genetic
        Stock Center (CGSC) database at Yale University, kindly supplied by
        Mary Berlyn. A public version of the database is accessible
        (http://cgsc.biology.yale.edu). Annotation of the genome is an
        ongoing task whose goal is to make the genome sequence more useful
        by correlating it with other data. Comments to the authors are
        appreciated. Updated information will be available at the E. coli
        Genome Project's World Wide Web site
        (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
        its annotations are periodically updated; this is version M54. No
        sequence changes. Annotation updates: updated gene identifications
        and products; all new functional assignments courtesy of Monica
        Riley; added promoters, protein binding sites, and repeated
        sequences described in reference 1. The unique numeric identifiers
        beginning with a lowercase 'b' assigned to each gene (protein- or
        RNA-encoding) are now designated as gene synonyms instead of
        labels. This should allow them to be searched for in Entrez as gene
        names.
FEATURES
    source
        1..11022
        /organism="Escherichia coli K12"
        /strain="K12"
        /sub_strain="MG1655"
        /db_xref="taxon:83333"
        <1..130
        /note="REP (repetitive extragenic palindromic) element;
        contains 4 REP sequences"
        /gene="ybhC"
        /note="b0772"
    repeat_region
        gene

```

```

KNEHEKMPYLEQWVQYKAMGLACMTLGTLSQAOALANAGLDYYNHNLDTSPEF
YGNITTRTYQERLDTLEKVRDAGIKVCSGCIIVGGETYKDRAGLLQILANIPPEP
VPLNMLVKYKGTADNDVDADFIRITAVNRIPMPTSYVRLSAGREOMNPTQAMC
PMAGANSIFYGCKLITTPNPEEKDLQFLRKILGNPGCAVLAGNEQOQRLQALMT
POTDEYNAAL"
RBS
2284. .2297
/standard_name="BIOC RBS"
gene
2295. .3050
/genes="BIOC"
CDS
2295. .3050
/genes="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKKQAIAAEGRAAHYEOHADLQROSDALLAMPKRYT
HYLDACGPGMWSRHWKERRHQAQVYALDLSPPMLYARQKDAADHYLAGDLESLPLATA
TDLAMNSLAVQWCGNLSTALRELYRVVRKGVAFPTLVQGLPELHQAMQAVDERP
HANRFLPDEIEQSLNGVYOHIIQPIITLMPDALSAMRSLKIGCATHLHEGRDRIL
TRSLQRLQALNQOQGRYPLTYHLGLVIARE"
3742. .3752
/standard_name="BIOA RBS"
gene
3750. .5039
/genes="BIOA"
CDS
3750. .5039
/genes="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/EC_start=1
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDQRIHIMPTYSMTSPLFVYPVSAEGCELLSDGR
LVGMSMAAITHGVNHPOLNAKMSIDAMSHVMGGITTHAPATELCRKLIVAMPPOP
LECVLADSGSAVAEYKAMKALQYQAKSEARQLTFPNCYHGTFCGMSYCDPDS
MISLKGCTLPENLFAPOSKMDGENDERDNGFALMAHREHETAAVITEPIVQAG
GMRWHPMLKRIKICDREGILLADELATGFRGKLFACENAEIADILICLQKAL
TGCTMTLSATLTREVAETISNGEGCFMHGPTFMQNLACAAANSLALIESGWOQ
OYADIEVQLREOLAPARDAENVADYVLCALGVETTHPVNMAALQKPFVEGVWIRP
FGKLIYLMPPYIILPQOLQRLTAANVAVQDETFQO"
5088. .5100
/standard_name="ORFI RBS"
gene
5098. .5574
/genes="ORFI"
CDS
5098. .5574
/genes="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDCKLPRHVPNGMGYGDGNISPHLAWDPVAGTK
SEVYCYDADAPGSGMHWVYVNLPAIDRVLPQGGSLVAMPDGVLTQTRDFSKTG
YDCAAPKGETHRIYFTVHALDIERIDVDEGASGAMVGFNVHFLSASLTVAMS"
5583. .5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

```

```

Qy 1 gacatctgcgaagtcacagaatta 25
|||||
Db 137 GACATCTGCCAAGTCACAGATTA 161
RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
REFERENCE
AUTHORS Birch, O. and Brass, J.
TITLES Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Source
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/genes="BIOF"
1154..2308
/genes="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRADALRRRYPVACAGRWLVADDRQYLN
SSNDYLGSHHPQIIRAMQOCAFQFGSGSGSVGHQALREBELAEMIGYSRA
LLPISGFANAVAIAMMAKEDRTAADRLSHASLLEASISPSOLREFANDVTHLAR
LLASPCQOQNVTEGYFMSNDGSAPLAETQVYQOQNGMLAMDADAGTYVEQGRG
SCWLOKVKPELLVVTGKFGVSGAAVLCSTVADYLLDFARHLITSTSPPAQOAL
RASLAVIRSDGDAAREKLAALLITRFAGVODLPFTLADSCSAIOPLVGDNSRALOL
AEKLRQOCQWVTAIRPTVPAGTARLRLTLAAHEMODIRLLLEVLHGNC"
3030..3045
/standard_name="BIOB RBS"
gene
3043..3753
/genes="BIOB"
CDS
3043..3753
/genes="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEYTGDTVEGTVASCALLQAKAAGRTAGYKRVASGS
EKTPGELNSDALALQNRNSSQLQDYAVAVNPYTAEPSPHIIISAQEGRPLESIVMSAG
LRALEQOQADWLVYDGAGGWEPLSDPTFPADWYTOEOLPVILVGVKLCGINHAMTLTA
OVIOHAGITLAGWVANDVTPPGKRHAHYMTTLTRMIPAPLLGELPWLAEENPENAATGK
YINLAFVADASTLGTFSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.043;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

stem_loop 5583. .5605
terminator 5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 137 GACATTGTGCGCAAGTCACAGAATTA 161

RESULT 7
A38251
LOCUS A38251 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:22994849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694* 960227.
location/Qualifiers
1. 5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141. .1156
/standard_name="BIOF RBS"
1154. .2308
/gene="BIOF"
1154. .2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKINALLDARRAADA.LRRRYVAVOAGAGNIVADDDROYLNF
SSNDYGLSHHPOITIRANQCAAFQEGTGGSGSHVSQSYVQVYHEELADMLGSKRA
LUTSGFAAMQAVIAAMAKEDRIADRI.SHASLLEAASI.SFSQJRRFMINDVTHLAR
LNASPCGQOMVTEGVFSMDSDSAPLAEIQOVTOQHNMVDAHGGVAGEGRC
SCWLOKRVKELLVTVKGFQVSGAAVILCSSTVADYLLQFARHLITSTMPAQOAL
RASLAVIRSDLEGVAREKLAALITRRACVODLPFLADSCAIDPLIYGDSPRALOL
AERLRQGCWCVTIRPTVPACTARLRLLTLAAHEMQDIDRLLEVIHGNG"
3030 3045
RBS /standard_name="BIOF RBS"
3043. .3753
gene /gene="BIOF"

CDS 3043. .3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEVTGDTVGVKTVASCALLQAKAAGTTRAGYKVASGS
EKTPGGLNSDALALQKNSSLQLDYAVVNPYTAEPSPHISAQEGRPLESIVMSAG
LRALQOQADWVTVGAGAGWFTPLSDTFEPADWYQEOPLVILVGVKICGINHAMLTA
OVIOHAGITLAGWAVANDVTPGCKRHAKEYMTTITRIPAPLIGELPWLAEKPEMAATGK
YINLAFVDSYRIGFTSRU"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 137 GACATTGTGCGCAAGTCACAGAATTA 161

RESULT 8
A93674
LOCUS A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
location/Qualifiers
1. 5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1. .96
/function="PROMOTER PTAC"
/evidence=experimental
23. .28
45. .49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105. .119
/standard_name="BIOF RBS NO.9"
/evidence=experimental
117. .1157
/gene="BIOF"
117. .1157
/gene="BIOF"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CA869590.1"
/db_xref="GI:6741863"
/translation="MAHRPRWTLISQVTELEFEKPLDLLPEAQOVHROHPDPOVYST
LLSIKTGACPEIDCKYCPQSSSKYKTLGLEAERLMEVEQVLESARAKAGSTRFCGAMW

[illegible]

```

ACCESSION      E00893
VERSION        E00893.1 GI:2169154
KEYWORDS       JP 1986149091-A/1.
SOURCE         Escherichia coli.
ORGANISM       Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
               Escherichia.
REFERENCE      1 (bases 1 to 1121)
AUTHORS       Hirono,Y., Kojima,T. and Kimura,H.
TITLE         DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
               PRODUCTION OF BIOTIN
JOURNAL       Patent: JP 1986149091-A 1 07-JUL-1986;
               NIPPON SODA CO LTD
COMMENT        OS Escherichia coli
               PN JP 1986149091-A/1
               PD 07-JUL-1986
               PF 24-DEC-1984 JP 1984272605
               PI HIRONO YOSHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PG
               C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18,PC
               C12R1:19);
               CC strandedness: Double;
               CC topology: Linear;
               CC hypothetical: No;
               CC anti-sense: No;
               CC *source: strain-Escherichia coli Nsl01;
               CC Feature is identified by experimental;
               FH key Location/Qualifiers
               FT CDS 42..1079
               /product="biotin synthetase"
               /location/Qualifiers
               source 1..1121
               /organism="Escherichia coli"
               /db_xref="taxon:562"
BASE COUNT     289 a 296 c 325 g 211 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacatgtcgaagtcacagaatta 25
   111111111111111111111111
Db 62 GACATTGTCCAGACGACGAATTA 86
RESUIT 5
ECOBIO 5793 bp DNA BCT 28-FEB-1994
LOCUS      E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
           protein, and dehydrobiotin synthetase (biob), complete cds.
ACCESSION  J04423
VERSION    J04423.1 GI:145422
KEYWORDS   7,8-diamino-pelargonic acid aminotransferase;
           7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
           bioc gene; biob gene; bioF gene; biotin synthetase; dehydrobiotin
           synthetase.
SOURCE     Escherichia coli (strain K-12) DNA.
ORGANISM   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (bases 1 to 5793)
AUTHORS   Otsuka,A.J., Buonocristiani,M.R., Howard,P.K., Flamm,J. and
           Johnson,O.
TITLE     The Escherichia coli biotin biosynthetic enzyme sequences
JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE   89066784
COMMENT   Draft entry and computer-readable sequence [1] kindly submitted by
           A. Otsuka, 09-NOV-1988.
FEATURES   Location/Qualifiers
           gene 1..5793
               /organism="Escherichia coli"
               /strain="K-12"
               /db_xref="taxon:562"
               /complement(98..574)
               /note="ORF 1"
               /codon_start=1
               /transl_table=1
               /protein_id="AAA23513.1"
               /db_xref="GI:451168"
               /translation="MKLISNDLRDGLPRHVFNGKYGDSNISPILADNDVPAGTK
               SFVVTGYDPAPTGSGMMHWVVALPADTETVLPOGFGSLVAMPDGLGTRTGGTG
               YDGAAPKRGETHRYIFTVHALDTERIDVDGASAAVGFVHSHLASASITAMPS"
               complement(633..1925)
               /gene="bioA"
               /complement(633..1925)
               /gene="bioA"
               /codon_start=1
               /transl_table=1
               /product="7,8-diamino-pelargonic acid aminotransferase"
               /protein_id="AAA23514.1"
               /db_xref="GI:457106"
               /translation="MTTDLAPDRHIMHPYTSMTSPLPYVSAEGCELILSDGR
               LVGSSWMAAIGHYNNPDLNAAMKSOIDAMSHVMFGITTHAPILCKRLVAMSARN
               ALDCLVPLADSGSAVEAYAMKALQYQAKGEARQFLTFRNGYHCDTFCGMSVCDPDN
               SMHSLMGYLPENLIFAPQSRMDGEDRDWGFARLMAAHRHEIAVATITPYGCA
               GKMRYHPEMLKRIKICDREGILLIADIEAFGRGKGLFACEHAELAPDLICLCKA
               LTGSTMTLSATLTITREVAETISNGEACGPRHGTTPGNPLACAAANASLITLESQWQ
               QGVADTIVOLREOLAPARDENYADVRLGALGVETFTTPVMAALQKFFVSGWIR
               FEGRLIYMPYITLPOQLRLTAANRAVODETFEQ"
               2012..3052
               /gene="bioB"
               /codon_start=1
               /transl_table=1
               /product="biotin synthetase"
               /protein_id="AAA23515.1"
               /db_xref="GI:145425"
               /translation="MAHRPRWTLISQVTELEFEKPLDILFEAOVROHRRDPPOVST
               LLSIKTGACPEDECKYCPOTSRYKTKLEARLMEVEQVYLESARAKAAGSTRCKMAAM
               KNHERDMPLYLEOMGVKAMGLEACNTIGTISBSQAOGLAAGIDTYINHLNLTSPER
               VPIINLVKVGTPLDNDVDVDAFETRTIAVARIMPPSVYLSAGRDQMNQOTQAMC
               EVAGANSIFYGCKLITTPNPEEDKDLQLEFRKLGLPDQTAVLACDNEQOQRLQALMT
               PDRDEYYNAAL"
               3049..4203
               /gene="bioF"
               /codon_start=1
               /transl_table=1
               /product="7-keto-8-amino pelargonic acid synthetase"
               /protein_id="AAA23516.1"
               /db_xref="GI:145426"
               /translation="MSWDEKINAALDARRADALRRYPVAGAGRWLVADROYLNF
               SSNDULGSHHPOIIRAWQQAEOFGISGSGSHVGYVYVHQALEELALMIGSRA
               LLEISGFANQAVIAAMAKEDRIADRIADLSHSLLEASLSPSQLRRFRAHNDVTHLAR
               LLASPCGGQNAVTEGVFSMDGSAPLAEIDQVTOHNGMLMVDDAHGTGVIIGBGRG
               SCMLQKVKPELVTFEGGFGVSAVACSTVADYILOFRHLITYSMPDAOAL
               RASLAVIRSDGEDDARRKRLALITFERAGYODIPTTLADSSAIOPLIVGNSRALOL
               AKRLQCCGWVAIRPPYVPAGTARLRLITLTAHEMODIDRLLEVLHNG"
               4190..4945
               /gene="bioC"
               /codon_start=1
               /transl_table=1
               /protein_id="AAA23517.1"
               /db_xref="GI:145427"
               /translation="WATVNRKQAIAAAFGRAAAHYEQHADLQROSADALLAMPORRYT
               HVLDGCGPQMSRHMERIRIAQVTLADISPEMLVQAROKDAADHYLAGIDTESPLATA
               TFDLWMSLAVQWCGNLSIALRELRYRVRPRGVAAFTTIVGSLPERHQAQNAVDERP

```

```

71 18.2 72.8 152118 9 AC008471 Homo sapi
72 18.2 72.8 154796 9 AC007372 Homo sapi
73 18.2 72.8 155633 2 AC025510 Homo sapi
74 18.2 72.8 155885 9 AL442646 human DNA
75 18.2 72.8 156766 9 AC079597 Homo sapi
76 18.2 72.8 160644 9 AC068770 Homo sapi
77 18.2 72.8 161391 9 AC021878 Homo sapi
78 18.2 72.8 161790 2 AC022371 Homo sapi
79 18.2 72.8 161812 9 HSI197017 Human DNA
80 18.2 72.8 164748 2 AC012139 Homo sapi
81 18.2 72.8 165046 2 AC091991 Homo sapi
82 18.2 72.8 165360 9 AC068725 Homo sapi
83 18.2 72.8 165523 2 AC022646 Homo sapi
84 18.2 72.8 168677 9 CNS0000F human chr
85 18.2 72.8 168749 2 AC008573 Homo sapi
86 18.2 72.8 169830 2 AC013345 Homo sapi
87 18.2 72.8 170048 2 AC025240 Homo sapi
88 18.2 72.8 172314 2 AC026154 Homo sapi
89 18.2 72.8 174479 2 AC092985 Homo sapi
90 18.2 72.8 175049 2 AC027047 Homo sapi
91 18.2 72.8 175757 2 AL356534 Human DNA
92 18.2 72.8 176040 2 AL583850 Homo sapi
93 18.2 72.8 177843 2 AC024446 Homo sapi
94 18.2 72.8 178568 9 HSJ1164C1 Human DNA
95 18.2 72.8 180157 2 AC021548 Homo sapi
96 18.2 72.8 181570 2 AC068738 Homo sapi
97 18.2 72.8 181632 2 AC053508 Homo sapi
98 18.2 72.8 182049 2 AC021239 Homo sapi
99 18.2 72.8 183118 2 AL355528 Homo sapi
100 18.2 72.8 183395 2 AC026042 Homo sapi

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Paton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    Location/Qualifiers
        source
            1..1041
            /organism="Unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gacattgtcgcaagtcacagaattta 25
    |||||||||||||||||||||||||
Db 21 GACATTGTGCGCAAGTCACAGAATTTA 45

```

```

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1041)
TITLE Paton,D.A.
JOURNAL Transgenic plants having increased biotin content
FEATURES
    source
        1..1041
        /organism="Unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gacattgtcgcaagtcacagaattta 25
    |||||||||||||||||||||||||
Db 21 GACATTGTGCGCAAGTCACAGAATTTA 45

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION Bior gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
FEATURES
    source
        1..1084
        /organism="Escherichia coli"
        /db_xref="taxon:562"
        /gene="B10B"
        /gene="B10B"
        /gene="B10B"
        /codon_start=1
        /transl_table=1
        /protein_id="CA00965.1"
        /db_xref="GI:490219"
        /db_xref="SMISS-PROT:P12996"
        /translation="MAHRPRWLTISOVLEFPEKPLDLLFPAQOVYHROHPPOVYST
        LLSITGACPEDCKYCCPOSSRYKGTGLAEARLMEYOEYVESARRKAKAGSTRFGAW
        KNPIERDPILEOVGVGKAMGLACACATLGLTSLSSOAKRLANGLDYNNINLPTSEF
        YGNITTRTKVOERLDLEKVDAGIKVCSGIVGLGTVKDRAGLLQIQLNLEPTPES
        VPIINMLYKGTPLADNDVDVAFDFITRTIATVARIIMPTSYVRISAGREQNEETOAMC
        FMAGANSIFYGCKLITTPNPEEDKDLQLFKRLGIMPQOTAVLAGDNQQRLRQALMT
        PDEYVMAAL"
        /protein_id="P12996"
        /db_xref="SMISS-PROT:P12996"
        /translation="MAHRPRWLTISOVLEFPEKPLDLLFPAQOVYHROHPPOVYST
        LLSITGACPEDCKYCCPOSSRYKGTGLAEARLMEYOEYVESARRKAKAGSTRFGAW
        KNPIERDPILEOVGVGKAMGLACACATLGLTSLSSOAKRLANGLDYNNINLPTSEF
        YGNITTRTKVOERLDLEKVDAGIKVCSGIVGLGTVKDRAGLLQIQLNLEPTPES
        VPIINMLYKGTPLADNDVDVAFDFITRTIATVARIIMPTSYVRISAGREQNEETOAMC
        FMAGANSIFYGCKLITTPNPEEDKDLQLFKRLGIMPQOTAVLAGDNQQRLRQALMT
        PDEYVMAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gacattgtcgcaagtcacagaattta 25
    |||||||||||||||||||||||||
Db 44 GACATTGTGCGCAAGTCACAGAATTTA 68

```

```

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin Synthetase.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:23:58 : Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-1

Sequence: 1 gacatgtcgaagtcacagaatla 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 100 summaries

Database : GenEmb1:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_hlg_hum:*
34: em_hlg_inv:*
35: em_hlg_rod:*
36: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	AL1530	AL1530 BioB gene
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	ECOB10	J04423 E.coli 7,8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A38251	A38251 Sequence 1
9	25	100.0	5872	6	A93674	A93674 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE000180	AE000180 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AF002553	AF002553 Escherich
15	23.4	93.6	5526	1	AF250776	AF250776 Unculture
16	21.6	87.2	8227	1	AF248314	AF248314 Unculture
17	21.4	85.6	17577	1	AC084797	AC084797 Homo sapi
18	20.2	80.8	965	1	AF250770	AF250770 Unculture
19	19.8	79.2	144355	9	AC016743	AC016743 Homo sapi
20	19.8	79.2	145495	2	AC093273	AC093273 Homo sapi
21	19.8	79.2	186735	2	AC093207	AC093207 Homo sapi
22	19.8	79.2	193661	2	AC090618	AC090618 Homo sapi
23	19.8	79.2	194576	2	AC019147	AC019147 Homo sapi
24	19.8	79.2	228434	2	AC012205	AC012205 Homo sapi
25	19.2	76.8	40897	3	CELC39D10	U39678 Caenorhabdl
26	19.2	76.8	107172	2	HS737M10	AL031388 Human DNA
27	19.2	76.8	146585	2	AC055761	AC055761 Homo sapi
28	19.2	76.8	149971	2	AC065937	AC065937 Homo sapi
29	19.2	76.8	160431	2	AL161796	AL161796 Homo sapi
30	19.2	76.8	167292	2	AC068365	AC068365 Homo sapi
31	19.2	76.8	173532	2	AC040890	AC040890 Homo sapi
32	19.2	76.8	180721	2	AC073466	AC073466 Homo sapi
33	19.2	76.8	181121	2	AC092895	AC092895 Homo sapi
34	19.2	76.8	190624	2	AC026892	AC026892 Homo sapi
35	19.2	76.8	197898	2	AC068781	AC068781 Homo sapi
36	19.2	76.8	204154	2	AL450352	AL450352 Homo sapi
37	19.2	76.8	233409	2	AC025126	AC025126 Homo sapi
38	18.8	75.2	32671	9	HS2A2A	Z84815 Human DNA
39	18.8	75.2	132444	9	AC019226	AC019226 Homo sapi
40	18.8	75.2	154669	2	AC076964	AC076964 Homo sapi
41	18.6	74.4	100184	9	AC008783	AC008783 Homo sapi
42	18.6	74.4	105207	9	AC011594	AC011594 Homo sapi
43	18.6	74.4	146002	9	CNS01RGD	AL157915 Human chr
44	18.6	74.4	155822	9	CNS01DIUV	AL133370 Human chr
45	18.6	74.4	161474	2	AC016198	AC016198 Homo sapi
46	18.6	74.4	179340	2	AC021065	AC021065 Homo sapi
47	18.6	74.4	349980	6	AX127150	AX127150 Sequence
48	18.2	72.8	4478	2	AC018106	AC018106 Drosophi
49	18.2	72.8	26058	3	AE003008	AE003008 Drosophi
50	18.2	72.8	62373	2	AC090045	AC090045 Homo sapi
51	18.2	72.8	78505	2	AC006408	AC006408 Homo sapi
52	18.2	72.8	80114	2	AC021542	AC021542 Homo sapi
53	18.2	72.8	86422	2	AC068672	AC068672 Homo sapi
54	18.2	72.8	92019	9	AL450320	AL450320 Human DNA
55	18.2	72.8	102204	9	AC003103	AC003103 Homo sapi
56	18.2	72.8	103952	9	AC008852	AC008852 Homo sapi
57	18.2	72.8	104117	9	AL591591	AL591591 Human DNA
58	18.2	72.8	117018	9	AC008790	AC008790 Homo sapi
59	18.2	72.8	123551	2	AC005809	AC005809 Homo sapi
60	18.2	72.8	134350	2	AC004905	AC004905 Homo sapi
61	18.2	72.8	135499	9	AC026446	AC026446 Homo sapi
62	18.2	72.8	139086	9	AL356127	AL356127 Human DNA
63	18.2	72.8	143712	9	HS917N8	AL031684 Human DNA
64	18.2	72.8	143967	2	AL357552	AL357552 Homo sapi
65	18.2	72.8	145246	2	AC087887	AC087887 Homo sapi
66	18.2	72.8	145342	2	AC023220	AC023220 Homo sapi
67	18.2	72.8	148498	2	AC034225	AC034225 Homo sapi
68	18.2	72.8	148573	2	AC066693	AC066693 Homo sapi
69	18.2	72.8	149440	2	AC024319	AC024319 Homo sapi
70	18.2	72.8	151974	2	AC027717	AC027717 Homo sapi

OY 3 atctgctgttgaagcgcagcag 25
|||||
DB 508 ATCTGCTGTTGATCTGCACGAG 530

RESULT 13
US-08-946-967-3
Sequence 3, Application US/08946967
Patent No. 5882869
GENERAL INFORMATION:
APPLICANT: Potter, Sharon L.
TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and
TITLE OF INVENTION: DNA Coding Therefor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,967
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1835 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 18..1469
OTHER INFORMATION: /product= "Maize Adenylosuccinate
US-08-946-967-3

Query Match 66.4%; Score 16.6; DB 2; Length 1835;
Best Local Similarity 82.6%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 atctgctgttgaagcgcagcag 25
|||||
DB 508 ATCTGCTGTTGATCTGCACGAG 530

RESULT 14
US-08-698-551-9/c
Sequence 9, Application US/08698551
Patent No. 5712381
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND

TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/698,551
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-698-551-9

Query Match 64.8%; Score 16.2; DB 1; Length 1694;
Best Local Similarity 85.7%; Pred. No. 11e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 5 ctgctgttgaagcgcagcag 25
|||||
DB 754 CTGCTGCCAGAGCCGACGAG 734

RESULT 15
US-08-602-228-9/c
Sequence 9, Application US/08602228
Patent No. 5843675
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/602,228
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-602-228-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgcgtgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 16
US-08-494-440B-9/c
Sequence 9, Application US/08494440B
Patent No. 5843501
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,440B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-494-440B-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgcgtgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 17
US-08-533-901B-9/c
Sequence 9, Application US/08533901B
Patent No. 5852173
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,901B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-533-901B-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgcgtgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 18
US-08-839-032A-9/c
Sequence 9, Application US/08839032A
Patent No. 5891675
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232DDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-839-032A-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 19
US-08-839-031A-9/c
Sequence 9, Application US/08839031A
Patent No. 5948638
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,031A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-839-031A-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 20
PCT-US95-12724-9/c
Sequence 9, Application PCT/US9512724
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12724
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
PCT-US95-12724-9

Query Match 64.8%; Score 16.2; DB 5; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcagcag 25
||||| |||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 21
US-07-755-573C-7
Sequence 7, Application US/07755573C
Patent No. 5622930

GENERAL INFORMATION:
APPLICANT: Eldering, Eric
APPLICANT: Arden, Lucien
TITLE OF INVENTION: C1 Esterase Inhibitor Mutelins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
STREET: 6900 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/755,573C
FILING DATE: 05-SEP-1991

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pochopien, Donald J.
REGISTRATION NUMBER: 32,167
REFERENCE/DOCKET NUMBER: 28687/32920
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1810 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1535
FEATURE:
NAME/KEY: mat_peptide

LOCATION: 102..1535
US-07-755-573C-7

Query Match 64.8%; Score 16.2; DB 1; Length 1810;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcagcag 25
||||| ||||||| |||||||
Db 1440 CTGCTGTTGAAGTCAGCAG 1460

RESULT 22
US-08-698-551-17/c
Sequence 17, Application US/08698551
Patent No. 5712381

GENERAL INFORMATION:
APPLICANT: Lin, Lin-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/698,551
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2473 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 14..2404
US-08-698-551-17

Query Match 64.8%; Score 16.2; DB 1; Length 2473;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcagcag 25
||||| ||||||| |||||||
Db 2227 CTGCTGCCAGAGCGCAGCAG 2207

RESULT 23
US-08-839-032A-17/c
Sequence 17, Application US/08839032A

```
Patent No. 5891675
GENERAL INFORMATION:
APPLICANT: Lin, Lin-Liang
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A., 32,724
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232DDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2473 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 14..2404
US-08-839-032A-17

Query Match 64.8%; Score 16.2; DB 2; Length 2473;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcacg 25
||||| |||||||||
Db 2227 CTGCTGCCAGAGCGCAGCAG 2207

RESULT 24
US-08-225-757B-1/C
Sequence 1, Application US/08225757B
Patent No. 5506133
GENERAL INFORMATION:
APPLICANT: YU, ET AL.
TITLE OF INVENTION: Superoxide Dismutase-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
```

```
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,757B
FILING DATE: 11 APR-94
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-106
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1080 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-225-757B-1

Query Match 64.0%; Score 16; DB 1; Length 1080;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgcgttttgaagcgcacga 24
||| ||||| ||||| |||
Db 385 GGGTCAGCTGTAGGAGCGCACCA 362

RESULT 25
US-08-722-050-1/C
Sequence 1, Application US/08722050
Patent No. 5871729
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: ROSEN, CRAIG A.
APPLICANT: FRASER, CLAIRE M.
APPLICANT: GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P. L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,050
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488,1020001/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
```

LENGTH: 1080 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (cdna)
FEATURE:
NAME/KEY: CDS
LOCATION: 115..879
US-08-722-050-1

Query Match 64.0%; Score 16; DB 2; Length 1080;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtcgttgaagcgacga 24
||| | ||||| ||||| ||
Db 385 GGGTCAGCTGTAGGAAGCGACCA 362

RESULT 26
US-08-910-505-3/c
Sequence 3, Application US/08910505A
Patent No. 6107071
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Raymond
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
APPLICANT: Lonetto, Michael
TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
FILE REFERENCE: P50549-02
CURRENT APPLICATION NUMBER: US/08/910,505A
CURRENT FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1248
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-08-910-505-3

Query Match 64.0%; Score 16; DB 3; Length 1248;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtcgttgaagcgacga 24
||| | ||||| ||||| ||
Db 968 GCATTACTGTTGAATCGACGCA 945

RESULT 27
US-08-910-505-1/c
Sequence 1, Application US/08910505A
Patent No. 6107071
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Raymond
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
APPLICANT: Lonetto, Michael

APPLICANT: Warren, Patrick
TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
FILE REFERENCE: P50549-02
CURRENT APPLICATION NUMBER: US/08/910,505A
CURRENT FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1251
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-08-910-505-1

Query Match 64.0%; Score 16; DB 3; Length 1251;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtcgttgaagcgacga 24
||| | ||||| ||||| ||
Db 968 GCATTACTGTTGAATCGACGCA 945

RESULT 28
US-08-176-620A-3/c
Sequence 3, Application US/08176620A
Patent No. 5595904
GENERAL INFORMATION:
APPLICANT: Boulton, Teri G.
APPLICANT: Cobb, Melanie H.
APPLICANT: Yancopoulos, George D.
APPLICANT: Nye, Steven
APPLICANT: Panayotatos, Nikos
TITLE OF INVENTION: A Family of Map2 Protein Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,620A
FILING DATE: 03-JAN-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1467 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1245
US-08-176-620A-3

Query Match 64.0%; Score 16; DB 1; Length 1467;

Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcagcagcag 25
| | | | | | | | | | | | | | | | |
Db 411 GTTCTCATGTCTGAAGCGCAGTAG 388

RESULT 29

US-08-463-862-3/C
; Sequence 3, Application US/08463862
; Patent No. 5776751
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G. et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,862
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,544
; FILING DATE: 16-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6526-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1467 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1245
; US-08-463-862-3

Query Match 64.0%; Score 16; DB 1; Length 1467;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcagcagcag 25
| | | | | | | | | | | | | | | | |
Db 411 GTTCTCATGTCTGAAGCGCAGTAG 388

RESULT 30

US-08-461-985-3/C
; Sequence 3, Application US/08461985
; Patent No. 5872006
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G.
; APPLICANT: Cobb, Melanie H.
; APPLICANT: Yancopoulos, George D.

APPLICANT: Nye, Steven
APPLICANT: Panayiotatos, Nikos
TITLE OF INVENTION: A Family of Map2 Protein Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,985
FILING DATE: 05-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/176,620
FILING DATE: 03-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1467 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1245
US-08-461-985-3

Query Match 64.0%; Score 16; DB 2; Length 1467;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcagcagcag 25
| | | | | | | | | | | | | | | | |
Db 411 GTTCTCATGTCTGAAGCGCAGTAG 388

RESULT 31

US-08-458-887-3/C
; Sequence 3, Application US/08458887
; Patent No. 5914261
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G. et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25


```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360,144A
FILING DATE: 20-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: APV-036.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 5430 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..5427
US-08-360-144A-11

Query Match          64.0%: Score 16; DB 3; Length 5430;
Best Local Similarity 79.2%: Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 ggaatcgctgttgaagcgacga 24
        || ||| ||||| || ||| |||
Db      3205 gattctgatgtttgtagtgtagca 3182

RESULT 39
PCT-US95-06722-11/c
; Sequence 11, Application PC/TUS9506722
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immunosuppressant Target Proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06722
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/250,795
; FILING DATE: 27-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/250,795
; FILING DATE: 20-DEC-1994
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97..7743
; PCT-US95-06722-11

Query Match          64.0%: Score 16; DB 5; Length 7824;
Best Local Similarity 79.2%: Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```


COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,656
FILING DATE: Concurrently Herewith
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/042,427
FILING DATE: 27-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/036,205
FILING DATE: 27-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/035,920
FILING DATE: 22-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hblber, David W.
REGISTRATION NUMBER: 41,071
REFERENCE/DOCKET NUMBER: UTSD:537
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ. ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 13865 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-009-656-11

Query Match 64.0%; Score 16; DB 3; Length 13865;
Best Local Similarity 79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 2 gatctgttgaagcgacagcag 25
|||||
Db 3657 GCTCTGCTTTTCGATCGACGAG 3634

RESULT 43
US-09-006-783A-6
Sequence 6, Application US/09006783A
Patent No. 6297366
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
APPLICANT: Garkavlev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
TITLE OF INVENTION: Pathway
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,783A
FILING DATE: 15-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6297366nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 633 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..630
US-09-006-783A-6

Query Match 63.2%; Score 15.8; DB 4; Length 633;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 7 gctgttgaagcgacagcag 25
|||||
Db 126 GCTGTTCGAGCGCGACGAG 144

RESULT 44
US-09-006-783A-4
Sequence 4, Application US/09006783A
Patent No. 6297366
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
APPLICANT: Garkavlev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
TITLE OF INVENTION: Pathway
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,783A
FILING DATE: 15-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6297366nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 873 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 7..813
US-09-006-783A-4

Query Match 63.2%; Score 15.8; DB 4; Length 873;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcgcagcag 25
 ||||| || ||||| |||||
 Db 339 GCTGTCGAGCGCGCAGCAG 357

RESULT 45

US-09-258-257-1
 : Sequence 1, Application US/09258257
 : Patent No. 5965398
 : GENERAL INFORMATION:
 : APPLICANT: GARKAVTSEV, Igor
 : APPLICANT: RIABOVOL, Karl
 : TITLE OF INVENTION: DNA SEQUENCE ENCODING A TUMOR
 : TITLE OF INVENTION: SUPPRESSOR GENE
 : NUMBER OF SEQUENCES: 12
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Burns, Doane, Swecker & Mathis
 : STREET: P.O. Box 1404
 : CITY: Alexandria
 : STATE: Virginia
 : COUNTRY: United States
 : ZIP: 22313-1404
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/258,257
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/569,721
 : FILING DATE: 08-DEC-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Mool, Leslie A.
 : REGISTRATION NUMBER: 37,047
 : REFERENCE/DOCKET NUMBER: 028722-128
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (650) 854-7400
 : TELEFAX: (650) 854-8275
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1902 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 109..738
 : US-09-258-257-1

Query Match

63.2%; Score 15.8; DB 2; Length 1902;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcgcagcag 25
 ||||| || ||||| |||||
 Db 234 GCTGTCGAGCGCGCAGCAG 252

Search completed: December 26, 2001, 12:23:07
 Job time: 7144 sec

86	17.6	70.4	728	11	BFS78750	BFS789250	602093285
87	17.6	70.4	755	13	AZ574713	AZ574713	305PVC05
88	17.6	70.4	741	13	CNS02KP2	AL201791	Tetraodon
89	17.6	70.4	753	11	BF1085414	BF1085414	602870097
90	17.6	70.4	820	11	BF583335	BF583335	602101611
91	17.6	70.4	851	11	B1258081	B1258081	602927279
92	17.6	70.4	883	11	BF2107222	BF2107222	601873580
93	17.6	70.4	890	11	BG391410	BG391410	602417459
94	17.6	70.4	917	11	BG767031	BG767031	602740565
95	17.6	70.4	931	11	BG744889	BG744889	602723244
96	17.6	70.4	946	13	CNS01IMO	AL145939	Anopheles
97	17.6	70.4	980	11	BF344943	BF344943	602014261
98	17.6	70.4	983	13	CNS02LPC	AL202737	Tetraodon
99	17.6	70.4	997	11	BG490046	BG490046	602518901
100	17.6	70.4	999	13	CNS03N7F	AL251700	Tetraodon

ALIGNMENTS

RESULT	1
A1514166	
LOCUS	A1514166 725 bp mRNA EST 23-Apr-2001
DEFINITION	GH72728.5prime GH drosophila melanogaster head pot2 Drosophila
ACCESSION	melanogaster cDNA clone GH72728.5, mRNA sequence.
VERSION	A1514166
KEYWORDS	A1514166.2 GI:13759256
SOURCE	EST.
ORGANISM	fruit fly, Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachytera; Mecoptomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 725)
REFERENCE	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
AUTHORS	BBCP/HMT Drosophila EST Project
TITLE	Unpublished (2001)
JOURNAL	On Mar 16, 1999 this sequence version replaced gi:4418228.
COMMENT	

FEATURES

son

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="CH27238"
/clone_lib="GH Drosophila melanogaster head por2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/notes="Organ: head; Vector: por2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
por2. Plasmid cDNA library."
BASE COUNT
164 a 210 c 205 g 146 t

```

	Query Match	Score	DB	Length
	Best Local Similarity	79.2%	19.8	725
	Matches	91.3%	Pred. No. 1.4e+02;	
	Conservative	0;	Mismatches	2;
			Indels	0;
			Gaps	0;
Qy	3	atctgctgtttaaagcagcagc	25	
Db	457	ATGTGCTGCTTGAAGCGCAGCAG	479	

RESULT	2
LOCUS	BEL14310/c
DEFINITION	BEL14310 507 bp mRNA EST 13-JUN-2000
ACCESSION	U1-R-BJ1-g-wk-g-02-0-U1.s1 U1-R-BJ1 Rattus norvegicus cDNA clone
VERSION	U1-R-BJ1-g-wk-g-02-0-U1.3, BEL14310 sequence.
KEYWORDS	BEL14310.1 GI:8506415
SOURCE	EST.
ORGANISM	Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 507) Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS	Normalization and subtraction: two approaches to facilitate gene
TITLE	discovey
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB

FEATURES

nos

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-awk-k9-02-0-UI"
/clone_lib="UI-R-BJ1"
/lab_host="MDH08 (Life Technologies)"
/notes="vector: pT73D fac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subcloned library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
clone.eng.uiowa.edu. The subtraction has been previously
described in (Bonardo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-BJ1
TAG_TISSUE=heart
TAG_SEQ=ACAAC

```

		77.6%;	Score 19.4;	DB 10;	Length 507;
Query Match	Similarity	95.2%;	Pred.	No. 2e+02;	
Matches	Conservative	0;	Mismatches	1;	Indels
					Gaps 0;
QY	4 lctgctgtltagagcgacga 24				
Dd	357 TCTGCTGTTTGAAGCGTCAGA 337				

Db 357 TCTGCTGTTGAAGCTCAGCA 337

RESULT 3
AA818607/c 528 bp mRNA EST 03-JUL-1999
LOCUS
DEFINITION U1-R-A0-bc-h-05-0-01.s1 U1-R-A0 Rattus norvegicus cDNA clone
U1-R-A0-bc-h-05-0-01.3' similar to gb|U96683|MM096683 Mus musculus
immunoglobulin-like receptor PIRAZ2 (6M23) mRNA, partial cds, mRNA
sequence.
ACCESSION AA818607 GI:4228398
VERSION
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 528)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Feb 17, 1998 this sequence version replaced gi:2889346.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult kidney library. cDNA library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.lnl.gov). IMAGE ID=1768272
Seq primer: M13 Forward
POLY-A-No.

FEATURES
source location/Qualifiers
1..528
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="U1-R-A0-bc-h-05-0-01"
/clone_lib="U1-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dr track which allows identification of
the library of origin of a clone within the mixture."

BASE COUNT 138 a 131 c 101 g 158 t
ORIGIN

Query Match 77.6%; Score 19.4; DB 10; Length 528;
Best local Similarity 95.2%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 tctgctgttgaaagcagca 24
|||||
Db 353 tctgctgttgaaagcagca 333

RESULT 4
LOCUS AO522448 437 bp DNA GSS 11-MAY-1999
DEFINITION HS-5237.A2.A12-T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-813 Col=24 Row=A, DNA sequence.
ACCESSION AO522448

VERSION AO522448.1 GI:4769482
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 437)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (<http://info@resgen.com>). BAC end web Server:
<http://www.htsc.washington.edu>
Plate: 813 row: A column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 437.

FEATURES
source location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=813 Col=24 Row=A"
/clone_lib="RPCI-11 Human Male BAC library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

BASE COUNT 101 a 125 c 112 g 93 t 6 others
ORIGIN

Query Match 76.8%; Score 19.2; DB 13; Length 437;
Best local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggarctgctgttgaaagcagca 24
|||||
Db 283 ggarctgctgttgaaagcagca 306

RESULT 5
LOCUS CNS054VD/c 1043 bp DNA GSS 26-JUL-2000
DEFINITION Tetradon nigroviridis genome survey sequence T3 end of clone
04J222 of library A from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL321250
VERSION AL321250.1 GI:9554134
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
Tetraodontidae; Tetradon.
1 (bases 1 to 1043)
Roest Croillius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot, A., Fizames, C., Wincker, P., Brothier, P., Queller, F., Saurin, W., and Weissensbach, J.
 Estimate of human gene number provided by genome-wide analysis using Tetradon nigroviridis DNA sequence
 Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL
REFERENCE
AUTHORS

TITLE
 2 (bases 1 to 1043)
 Crotti, H.R., Jallion, O., Dastiv, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Queller, F., Saurin, W., Bernot, A., and Weissensbach, J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetradon nigroviridis
 Genome Res. 10 (7), 935-949 (2000)

JOURNAL
REFERENCE
AUTHORS

TITLE
 3 (bases 1 to 1043)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetradon>.
 Location/Qualifiers
 1..1043
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone_id="A"
 /note="Genoscope sequence ID : C0MA043DEJ1A1-end : T3"
 BASE COUNT
 277 a 234 c 249 g 267 t 16 others

Query Match
 Best Local Similarity 87.5% Pred. No. 2.6e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatcgtcttgaagcagcag 25
 ||||| ||||| ||||| |||||
 Db 849 GATCTGCCGTTTAAAGCAGACAG 826

RESULT 6
LOCUS AV231852 238 bp mRNA EST 03-NOV-1999
DEFINITION AV231852 RIKEN full-length enriched, 0 day neonate skin Mus musculus cDNA clone 4632410E09 3', mRNA sequence.
ACCESSION AV231852
VERSION AV231852.1 GI:6184367
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
AUTHORS

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Teleno, M., Tomihata, N., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al. 1999)
 Unpublished (1999)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Saitama-cho, Tsukuba-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome_research@riken.go.jp
 URL: <http://genome.gsc.riken.go.jp/>
 Sasaki, N., Izawa, M., Watanabe, M., Okazaki, Y., Yoneda, Y., Matsuyama, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itou, M., Kikuchi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P., and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (<http://genome.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
 1..238
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_id="4632410E09"
 /clone_lib="RIKEN full-length enriched, 0 day neonate skin"
 /sex="mixed"
 /tissue_type="skin"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGGATTCGAGTTAAATTAATCCCCCCCCCC 3'1. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC19 after bulk excision from Lambda FLX I"

BASE COUNT

67 a 59 c 61 t

Query Match

Best Local Similarity 90.9% Pred. No. 3.3e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tctgctgttgaagcagcag 25
 ||||| ||||| ||||| |||||
 Db 88 TCTGCTGTTTAAATCAGACAG 109

RESULT 7
LOCUS BG726267 470 bp mRNA EST 09-MAY-2001
DEFINITION BG726267 sra13809.y1 gm-cl067 glycine max cDNA clone GENOME SYSTEMS CLONE
ACCESSION BG726267
VERSION BG726267.1 GI:14011345
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolaeae;

REFERENCE 1 (bases 1 to 470)
 Glycine.
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvill, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 411.
 Location/Qualifiers
 1..470
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl067-2993"
 /clone_1ib="Gm-cl067"
 /tissue_type="Germinating shoot, 3 day old seedling, auxin treatment"
 /lab_host="DH10B"
 /note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from germinating shoots of 3 day old seedling for the cultivar Williams 82. The seedlings were germinated in a growth chamber using germination paper in a solution containing 100ppm auxin. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 130 a 82 c 101 g 157 t

ORIGIN

Query Match 75.2%; Score 18.8; DB 11; Length 470;
 Best Local Similarity 90.9%; Pred. No. 3.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tctgctgttgaagcagcagc 25
 ||||| ||||| ||||| |||||
 Db 428 TCTGCTGTTGAAGTCAGCAGC 449

RESULT 8
 AM944633 520 bp mRNA EST 05-JUL-2001
 LOCUS SD10401.3orfme SD Drosophila melanogaster Schneider I2 cell culture
 DEFINITION SD10401.3orfme SD Drosophila melanogaster cDNA clone SD10401.3 similar to
 AJ243916: Drosophila melanogaster Domina gene, exons 1-3, mRNA sequence.
 ACCESSION AM944633.1 GI:8122381
 VERSION AM944633
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE 1 (bases 1 to 520)
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.
 BDGP/HMT Drosophila EST Project
 Unpublished (2001)
 Other ESTs: SD10401.5orfme
 Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated. The resulting Poly-T sequence has been removed. hit genomic AF001686: Drosophila melanogaster genomic scaffold 14200013386035 section 11 of 105, complete sequence.. 03/20/2001
 Plate: SD.104 row: A column: 1
 High quality sequence stop: 474.
 Location/Qualifiers
 1..520
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="SD10401"
 /clone_1ib="SD Drosophila melanogaster Schneider I2 cell culture por2"
 /lab_host="DH5-alpha"
 /note="Vector: por2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into por2. Plasmid cDNA library."

BASE COUNT 152 a 100 c 117 g 150 t 1 others

ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 520;
 Best Local Similarity 90.9%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcagc 23
 ||||| ||||| ||||| |||||
 Db 340 GATCTGCTCTTGAAGCGCAGC 361

RESULT 9
 AM397892 575 bp mRNA EST 07-FEB-2000
 LOCUS S696e12.y1 Gm-cl1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-cl1007-1703 5' similar to TR004471 004471 SIMILAR TO
 SACHAROMYCES HYPOTHEICAL PROTEIN P9642.2. ; mRNA sequence.
 ACCESSION AM397892
 VERSION AM397892.1 GI:6916362
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae; Glycine.

REFERENCE 1 (bases 1 to 575)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvill, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Glibco
 High quality sequence stop: 416.
 Location/Qualifiers
 1..575
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl007-1703"
 /clone_1id="Gm-cl007"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
 cDNA library was constructed from mRNA isolated from
 immature cotyledons of greenhouse grown plants (individual
 seed fresh weight of 100-300mg). The library was prepared
 using the Life Technologies SuperScript cDNA library
 construction kit. Complementary DNA was synthesized from
 mRNA using a poly (dT) sequence with a NotI restriction
 site. SalI linkers adapters were ligated to the
 blunt-ended cDNA fragments followed by NotI digestion. The
 cDNA fragments were directionally cloned into the
 NotI-SalI restriction site of the pSPORT1 vector. The
 ligated cDNA fragments were transformed into E. coli
 Electromax DH10B host cells. This library was constructed
 by Dr. Lilla Voklin and Dr. Anu Khanna."

BASE COUNT 185 a 104 c 120 g 166 t
 ORIGIN

Query Match 75.2% Score 18.8; DB 10; Length 575;
 Best Local Similarity 90.9%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tctctgttgaagcgcagcag 25
 ||| ||||| ||||| |||||
 Db 117 TCTGCTGTTTGACGACGACG 138

RESULT 10
 CNS02Y61/c
 LOCUS
 DEFINITION
 180C23 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL219267.1 GI:7878086
 VERSION
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 802)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 802)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brotier, P., Quetier, F.,
 Sautin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 802)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Genoscope.
 Direct Submission
 Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 1..802
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="180C23"
 /clone_1id="G"
 /note="Genoscope sequence ID : C0AG180AB12LP1-end : 77"

BASE COUNT 187 a 166 c 188 g 254 t
 ORIGIN

Query Match 75.2% Score 18.8; DB 13; Length 802;
 Best Local Similarity 83.3%; Pred. No. 3.7e+02;
 Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatcgtcttgaagcgcagcag 25
 ||| ||||| ||||| |||||
 Db 668 GWTGCTGTTTGACGACGACG 645

RESULT 11
 AJ394746
 LOCUS
 DEFINITION
 AJ394746
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 Gallus gallus
 chicken.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 446)
 Abdurkhanov, I., Lodgyn, D., Gerolt, P., Arakawa, H., Law, A., Plachy
 J., Korn, B. and Buerstedde, J.M.
 A large database of chickenursal ESTs as a resource for the
 analysis of vertebrate gene function
 Genome Res. 10 (12), 2062-2069 (2000)
 Contact: Buerstedde JM
 Cellular Immunology
 Heinrich-Pette-Institute
 Martinistr. 52, 20251 Hamburg, Germany
 Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
 Location/Qualifiers
 1..446
 /organism="Gallus gallus"
 /strain="CB"
 /db_xref="taxon:9031"
 /clone="1119r1"
 /clone_1id="dkfz426"
 /tissue-type="fursa of Fabricsius"

JOURNAL
 MEDLINE
 COMMENT
 TITLE
 JOURNAL
 REFERENCE

FEATURES
 SOURCE

BASE COUNT 122 a 107 c 117 g 100 t
 ORIGIN

Query Match 74.4% Score 18.6; DB 10; Length 446;
 Best Local Similarity 84.0%; Pred. No. 4.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggatcgtcttgaagcgcagcag 25
 ||| ||||| ||||| |||||
 Db 407 GGACTGCTGTTGACGACGACG 431

RESULT 12

AV680177 468 bp mRNA EST 05-OCT-2000
LOCUS
DEFINITION AV680177 Nori Satoh unpublished cDNA library Clona intestinalis
CDNA clone rcitbln19 3', mRNA sequence.
AV680177
ACCESSION
VERSION AV680177.1 GI:10118176
KEYWORDS
SOURCE EST.
ORGANISM Clona intestinalis.
Clona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Cloniidae; Clona.
1 (bases 1 to 468)
REFERENCE Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
AUTHORS Expressed genes in Clona intestinalis
TITLE Unpublished (2000)
JOURNAL Contact: Nori Satoh
COMMENT Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
FEATURES
SOURCE
1. 468
/organism="Clona intestinalis"
/db_xref="taxon:7719"
/clone="rcitbln19"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"
/dev_stage="tailbud"
BASE COUNT 107 a 136 c 71 g 154 t
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 468;
Best Local Similarity 84.0%; Pred. No. 4.3e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 270 GGATCTGCTGTTGCAAGCTAAACAG 294
RESULT 13
AV677254 509 bp mRNA EST 05-OCT-2000
LOCUS
DEFINITION AV677254 Nori Satoh unpublished cDNA library Clona intestinalis
CDNA clone rcitb59 3', mRNA sequence.
AV677254
ACCESSION
VERSION AV677254.1 GI:10115253
KEYWORDS
SOURCE EST.
ORGANISM Clona intestinalis.
Clona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Cloniidae; Clona.
1 (bases 1 to 509)
REFERENCE Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
AUTHORS Expressed genes in Clona intestinalis
TITLE Unpublished (2000)
JOURNAL Contact: Nori Satoh
COMMENT Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
FEATURES
SOURCE
1. 509
/organism="Clona intestinalis"
/db_xref="taxon:7719"
/clone="rcitb59"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"

BASE COUNT 112 a 147 c 88 g 162 t
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 509;
Best Local Similarity 84.0%; Pred. No. 4.3e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 198 GGATCTGCTGTTGCAAGCTAAACAG 222
RESULT 14
AV675262 617 bp mRNA EST 05-OCT-2000
LOCUS
DEFINITION AV675262 Nori Satoh unpublished cDNA library Clona intestinalis
CDNA clone citbln19 5', mRNA sequence.
AV675262
ACCESSION
VERSION AV675262.1 GI:10113261
KEYWORDS
SOURCE EST.
ORGANISM Clona intestinalis.
Clona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Cloniidae; Clona.
1 (bases 1 to 617)
REFERENCE Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
AUTHORS Expressed genes in Clona intestinalis
TITLE Unpublished (2000)
JOURNAL Contact: Nori Satoh
COMMENT Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
FEATURES
SOURCE
1. 617
/organism="Clona intestinalis"
/db_xref="taxon:7719"
/clone="citbln19"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"
/dev_stage="tailbud"
BASE COUNT 187 a 130 c 169 g 130 t 1 others
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 617;
Best Local Similarity 84.0%; Pred. No. 4.4e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 502 GGATCTGCTGTTGCAAGCTAAACAG 478
RESULT 15
B1144862 628 bp mRNA EST 05-JUL-2001
LOCUS
DEFINITION B1144862 602909701F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5050657 5',
mRNA sequence.
B1144862
ACCESSION
VERSION B1144862.1 GI:14604863
KEYWORDS
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 628)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

1. 904
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COB6095CC06SPL-end :
 PUC-ori"

BASE COUNT 231 a 201 c 220 g 251 t 1 others
 ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 904;
 Best Local Similarity 84.0%; Pred. No. 4.6e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcagcagc 25
 ||||||||||| 1 ||||| 1
 db 626 GGATCTGCTGTTGACCTCAGCG 602

RESULT 18

CNS03FNU 930 bp DNA GSS 17-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
 DEFINITION 02J13 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL241912
 AL241912.1 GI:7962681
 GSS: genome survey sequence.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 930)
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis

JOURNAL 2 (bases 1 to 930)
 Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

Best Local Similarity 84.0%; Pred. No. 4.6e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcagcagc 25
 ||||||||||| 1 ||||| 1
 db 781 GGACCTGCTGTTGAAAAGACGACG 805

RESULT 19

CNS051ED 610 bp DNA GSS 26-JUL-2000
 LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 051011 of library A from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL316750
 AL316750.1 GI:9549634
 GSS: genome survey sequence.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

JOURNAL

AUTHORS

COMMENT

VERSION AV142481.1 GI:5346476
 KEYWORDS EST.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 185)
 AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Alizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nishitani, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomioka, N., Muramatsu, M., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs
 JOURNAL Unpublished (1999)
 COMMENT Contact: Chie Owa
 Genome Science Laboratory
 RIKEN
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@rctc.riken.go.jp
 Thermostabilization and thermocyclization of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
 Please visit our web site (http://genome.rctc.riken.go.jp) for further details.

FEATURES
 Location/Qualifiers
 1..185
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2810423K17"
 /clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
 /sex="mixed"
 /dev_stage="10-11 day embryo"
 /dev_stage="10-11 day embryo"
 46 c 38 g 50 t

BASE COUNT
 51 a 46 c 38 g 50 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 185;
 Best Local Similarity 87.0%; Pred. No. 5.8e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctgctgttgaaagcagcag 25
 ||||| ||||| ||||| |||||
 DB 41 ATCTGCTTTTGATCAGCAG 63

RESULT 21
 AA750666 243 bp mRNA EST 20-JAN-1998
 LOCUS AA750666
 DEFINITION IGS0248 Rice Immature Seed Lambda ZAPIT cDNA Library Oryza sativa
 ACCESSION CDNA clone IGS0248, mRNA sequence.
 VERSION AA750666.1 GI:2797372
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiaceae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 243)
 AUTHORS Nahn, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P., Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.T., Kang, K.Y., Lee, M.C. and Eun, M.Y.
 Large scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)
 COMMENT Contact: Eun M.Y.

Department of Cyogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@un20.asli.re.kr
 Submitted by Baek Hie Nahn, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@loserver.myongji.ac.kr
 Seq primer: M3 Reverse primer.

FEATURES
 Location/Qualifiers
 1..243
 /organism="Oryza sativa"
 /cultivar="Milyang23"
 /db_xref="taxon:4530"
 /clone="IGS0248"
 /clone_lib="Rice Immature Seed Lambda ZAPIT cDNA Library"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /note="Vector: pBluescript SK(+). Site 1: EcoRI; Site 2: XhoI; Directional cDNA library inserted into lambda ZAPIT vector at 5' end with EcoRI and 3' end with Xho I site."

BASE COUNT
 65 a 60 c 62 g 56 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 243;
 Best Local Similarity 87.0%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgttgaaagcagcag 24
 ||||| ||||| ||||| |||||
 DB 240 GATCTGCTGTGTGAGCTCACC 218

RESULT 22
 AV282092 251 bp mRNA EST 08-NOV-1999
 LOCUS AV282092
 DEFINITION musculus cDNA clone 493428B13 3', mRNA sequence.
 ACCESSION AV282092
 VERSION AV282092.1 GI:6274841
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 251)
 AUTHORS Kono, H., Alizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, T., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Kono, H., et al. 1999)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@rctc.riken.go.jp
 URL: http://genome.gsc.riken.go.jp/
 Sasaki, N., Izawa, M., Watanabe, M., Watanabe, M., Okazaki, Y., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.
 Location/Qualifiers

FEATURES
source

1. 251
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4933428B13"
 /clone_1lb="RIKEN full-length enriched, adult male testis
 (DH10B)"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGATTCGAGTTATTAATTATCCCCCCCCCC 3']. cDNA
 was cloned into the XhoI and BamHI sites. Vector: a
 modified plusscript KS(+) after bulk excision from Lambda
 FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
 62 a 57 c 57 g 75 t

BASE COUNT
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 251;
 Best Local Similarity 87.0%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctgctgttgaagcgacagcag 25
 ||||| ||||| |||||
 Db 104 ATCTGCTTTTGAATCAGCAG 126

RESULT 23
 BB452775/c mRNA EST 21-OUT-2000
 LOCUS BB452775 RIKEN full-length enriched, 12 days embryo spinal ganglion
 DEFINITION Mus musculus cDNA clone D130019M02.3' similar to X76772 M.musculus
 mRNA for ribosomal protein S3, mRNA sequence.

ACCESSION BB452775
 VERSION BB452775.1 GI:9348223
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 268)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Atakawa, T., Carninci,
 H., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata

TITLE
 JOURNAL
 COMMENT

Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya,
 T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamana, I.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
 M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagacka, S., Sasaki,
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermo-stabilization and thermo-activation of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES
source

1. 268
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D130019M02"
 /clone_1lb="RIKEN full-length enriched, 12 days embryo
 spinal ganglion"
 /tissue_type="spinal ganglion"
 /dev_stage="12 days embryo"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGCGCCGACGCTTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGATTCGAGTTATTAATTATCCCCCCCCCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 plusscript KS(+) after bulk excision from Lambda FLC I."
 80 a 71 c 57 g 60 t

BASE COUNT
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 268;
 Best Local Similarity 87.0%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacagc 24
 ||||| ||||| |||||
 Db 28 GATCTGCACTTTGAAGCAGCA 6

RESULT 24
 AV216567/c mRNA EST 30-OCT-1999
 LOCUS AV216567 RIKEN full-length enriched, ES cells Mus musculus cDNA
 DEFINITION

ACCESSION AV422424
 VERSION AV422424.1 GI:7777272
 KEYWORDS EST.
 SOURCE Lotus japonicus.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 481)
 ASAMIZU, E., NAKAMURA, Y., SATO, S. and TABATA, S.
 Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus
 DNA Res. 7 (2), 127-130 (2000)

JOURNAL MEDLINE
 20277479
 CONTACT: Yasukazu Nakamura
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 Location/Qualifiers

FEATURES
 source
 1..481
 /organism="Lotus japonicus"
 /db_xref="taxon:34305"
 /clone="MM013b10.r"
 /dev_stage="young plants (two-week old)"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; Isolate: Miyajima MG-20"

BASE COUNT
 104 a 113 c 115 g 149 t

ORIGIN
 Query Match 72.8%; Score 18.2; DB 10; Length 481;
 Best Local Similarity 87.0%; Pred. No. 6.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctgtcttgaagcgcagc 25
 |||||
 Db 240 ATCTGCTTTGGAGCCTCATCAG 218

RESULT 29
 BG160914/c 584 bp mRNA EST 06-FEB-2001
 LOCUS de38c07.x1 Wellcome CRC PRN3 dorsal lip Xenopus laevis cDNA clone
 DEFINITION IMAGE:3474037 3' similar to TR:Q15059 Q15055 KIAA0043 PROTEIN. ;,
 mRNA sequence.

ACCESSION BG160914
 VERSION BG160914.1 GI:12694833
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 584)
 CLIFTON, S., JOHNSON, S.L., BLUMBERG, B., SONG, J., HILLIER, L., PAPE, D., MARTIN, J., WYLIE, T., UNDERWOOD, K., THEISING, B., BOWERS, Y., PERSON, B., GIBBONS, M., HARVEY, N., RITTER, E., JACKSON, Y., MCCANN, R., WATERSTON, R. and WILSON, R.
 WashU Xenopus EST project, 1999
 Unpublished (1999)

TITLE
 JOURNAL
 COMMENT
 Contact: Sandy Clifton, Ph.D.
 WashU Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by A.M. Zorn (Wellcome/7Kc Institute). DNA
 Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LNM at: info@image.llnl.gov
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40UP from Gibco
 High quality sequence stop: 494.
 Location/Qualifiers

FEATURES
 source
 1..584
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3474037"
 /clone_lib="Wellcome CRC PRN3 dorsal lip"
 /tissue_type="dorsal lip"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library was constructed by A.M. Zorn (Wellcome/CRC Institute)."

BASE COUNT
 227 a 108 c 163 g 86 t

ORIGIN
 Query Match 72.8%; Score 18.2; DB 11; Length 584;
 Best Local Similarity 87.0%; Pred. No. 6.6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggatctgtcttgaagcgcagc 23
 |||||
 Db 470 GGATTGCTGCTTTGGCGCCGACG 448

RESULT 30
 A1325366/c 685 bp mRNA EST 23-DEC-1998
 LOCUS m129f01.y1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA
 DEFINITION clone IMAGE:464953 5' similar to TR:060947 060947 MAX-INTERACTING
 TRANSCRIPTIONAL REPRESSOR. ;, mRNA sequence.

ACCESSION A1325366
 VERSION A1325366.1 GI:4059795
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 685)
 MARRA, M., HILLIER, L., ALLEN, M., BOWLES, M., DIETRICH, N., DUBUQUE, T., GEISEL, S., KUCABA, T., LACY, M., LE, M., MARTIN, J., MORRIS, M., SCHELLENBERG, K., SEPTOE, M., TAN, F., UNDERWOOD, K., MOORE, B., THEISING, B., WYLIE, T., LENNON, G., SOARES, B., WILSON, R. and WATERSTON, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)

TITLE
 JOURNAL
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouse@est.wustl.edu
 This clone is available royalty-free through LNM; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:278769
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 correct orientation)
 This read has been verified (found to hit its original self in the
 possible reversed clone: similarity on wrong strand
 seq primer: -40RP from Gibco
 High quality sequence stop: 384.
 Location/Qualifiers

FEATURES
 source
 1..685
 /organism="Mus musculus"
 /strain="C57Bl/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:464953"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"

```

/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAAGTGGAGCGCGGCAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonalido."

```

BASE COUNT 158 a 194 c 230 g 103 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 685;
 Best Local Similarity 87.0%; Pred. No. 6.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggaatcgtgtttgaagcagc 23
 |||||||||||||||||

DB 455 ggctgctgtttgaagcagc 433

RESULT 31

LOCUS AU170196 765 bp mRNA EST 29-JAN-2001

DEFINITION AU170196 Ol-br-ad cDNA Oryzias latipes cDNA clone br5759, mRNA sequence.

ACCESSION AU170196

VERSION AU170196.1 GI:12592267

KEYWORDS EST.

SOURCE

ORGANISM Japanese medaka.

Oryzias latipes.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

1 (bases 1 to 765)

Mita,K., Ishikawa,Y. and Yamauchi,M.
 Establishment of cDNA database of medaka, Oryzias latipes
 Unpublished (2001)
 Contact: Mita K
 Genome Research Group
 National Institute of Radiological Sciences
 Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
 Email: kmitsa@irs.go.jp
 method:un1-directional sequence direction:sequenced from T3 primer
 (5' -> 3').

FEATURES

source

1..765

Location/Qualifiers

/organism="Oryzias latipes"

/strain="HNI"

/db_xref="taxon:8090"

/clone_lib="br5759"

/clone_lib="Ol-br-ad cDNA"

/sex="female/male mixed"

/tissue_type="brain"

/dev_stage="adult"

BASE COUNT 238 a 157 c 146 g 224 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 765;
 Best Local Similarity 87.0%; Pred. No. 6.8e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctgctgttgaagcagcag 25

```

|||||
DB 449 AGCTGCTGTTTGAAGCGCAGCAG 471

```

RESULT 32

LOCUS BG107933/C 931 bp mRNA EST 30-JAN-2001

DEFINITION BG107933 602278215p1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4365890 5', mRNA sequence.

ACCESSION BG107933

VERSION BG107933.1 GI:12601779

KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 931)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA library preparation: Life Technologies, Inc.

cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10015 row: n column: 03.

FEATURES

source

1..931

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:4365890"

/clone_lib="NIH_MGC_86"

/tissue_type="osteosarcoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.53 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC library."

BASE COUNT 279 a 222 c 213 g 216 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 931;
 Best Local Similarity 87.0%; Pred. No. 6.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggaatcgtgtttgaagcagc 23
 ||| || ||||||| |||||||

DB 69 GGACTGCTGTTTGAAGCGCAGC 47

RESULT 33

LOCUS BG888762/C 381 bp mRNA EST 30-MAY-2001

DEFINITION EST514613 cSTD Solanum tuberosum cDNA clone cSTD11014 5' sequence, mRNA sequence.

ACCESSION BG888762

VERSION BG888762.1 GI:14265848

KEYWORDS EST.

SOURCE

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 381)

van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemlingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.

Generations of ESTs from dormant potato tubers

JOURNAL COMMENT

Unpublished (2001)
Contact: Cathy Konning
The Institute for Genomic Research
For clone info: please contact Research Genetics Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13P-R.

FEATURES

Location/Qualifiers
1..381

/organism="Solanium tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cspdl1014"
/clone_lib="cspd"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tubers, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 4°C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."

BASE COUNT

88 a 76 c 111 g 106 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 381;
Best Local Similarity 90.5%; Pred. No. 9.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

5 ctgctgtttaagcgacgacg 25
|||
Db 315 CTCTTGTTCGAGCGCAGCAG 295

RESULT 34

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

LOCUS: A0464763 525 bp DNA GSS 23-APR-1999
DEFINITION: HS-5105_B2-B05-T7A RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION: A0464763
VERSION: A0464763.1 GI:4641858
KEYWORDS: GSS.
SOURCE: human.
ORGANISM: Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE: 1 (bases 1 to 525)
Mahalras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
CONTACT: Mahalras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research Genetics (info@resgen.com). BAC end web Server: http://www.htsc.washington.edu
Plate: 681 row: D column: 10
Seq primer: T7
Class: BAC ends

FEATURES

source

High quality sequence stop: 525.

Location/Qualifiers
1..525

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pJate-681 Col=10 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT

123 a 131 c 99 g 166 t 6 others

ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 525;
Best Local Similarity 90.5%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

4 tctgctgtttaagcgacgacg 24
|
Db 269 TGTGCTGTTTGAAGCCCGACGA 249

RESULT 35

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

LOCUS: AM644558 534 bp mRNA EST 26-APR-2001
DEFINITION: cm4e03.w1 Blackshear/Soares normalized Xenopus egg library Xenopus laevis cDNA clone PBX0141E03 5', mRNA sequence.
ACCESSION: AM644558
VERSION: AM644558.1 GI:7401949
KEYWORDS: EST.
SOURCE: African clawed frog.
ORGANISM: Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopus.
REFERENCE: 1 (bases 1 to 534)
Blackshear,P.J., Lai,W.S., Thonn,J.M., Kennington,E.A., Staffa,N.G., Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman,J.W., Bonaldo,M.F. and Soares,M.B.
The NIHHS Xenopus maternal EST project: Interim analysis of the first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
CONTACT: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709, USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 alt:cdna, email cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCCAGT
BACKWARD: CAGGAACGACGATGACG
Plate: 0141 row: E column: 03
Seq primer: T7 primer.
Location/Qualifiers
1..534
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0141E03"
/clone_lib="Blackshear/Soares normalized Xenopus egg library"

```

/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: p773-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
p773-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery'. Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested p773-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."

```

```

BASE COUNT      138 a      125 c      149 g      122 t
ORIGIN

```

```

Query Match      71.2%; Score 17.8; DB 10; Length 534;
Best Local Similarity 90.5%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY      3 atctgctgttgaagcagcagc 23
      ||||| ||||| ||||| |||||
Db      283 ATCTCCTTTGAAAGTGCAGCAGC 303

```

```

RESULT 36
LOCUS   A1676750      553 bp      mRNA      EST      19-MAY-1999
DEFINITION
etmST0529 Eth1 Eimeria tenella cDNA clone etmj093 5', mRNA
sequence.
ACCESSION A1676750
VERSION   A1676750.1 GI:4877230
KEYWORDS EST.
SOURCE   Eimeria tenella.
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.

```

```

REFERENCE 1 (bases 1 to 553)
AUTHORS   Wan, K.L., Chong, S.P., Ng, S.T., Tomley, F.M. and Jang, M.S.
TITLE     Survey of genes in Eimeria tenella merozoite
JOURNAL   Unpublished (1999)
COMMENT   Contact: Wan K.L.
Center for Gene Analysis and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor DE, Malaysia
Tel: 6 03 8293997
Fax: 6 03 8293249
Email: klwan@pkriscc.cc.ukm.my
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: SK

```

```

FEATURES
source
Location/Qualifiers
1..553
/organism="Eimeria tenella"
/strain="Houghton"
/db_xref="taxon:5802"
/clone="etmj093"
/clone_lib="Eth1"
/dev_stage="second generation merozoite"
/lab_host="XLI-Blue MRF"
/notes="Vector: Lambda ZAP1; Site_1: EcoRI; Site_2: XhoI;
Second generation merozoites of E. tenella H were purified
by column chromatography and mRNA was synthesized using a
FASTTRACK kit (Invitrogen). cDNA was synthesized and a
Uni-ZAP XR library was constructed using cDNA synthesis
kit, ZAP-cDNA synthesis kit and ZAP-cDNA GigaPack III
Gold cloning kit (Stratagene). The library was amplified

```

```

BASE COUNT      133 a      143 c      164 g      112 t      1 others
ORIGIN

```

```

Query Match      71.2%; Score 17.8; DB 10; Length 553;
Best Local Similarity 90.5%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY      5 ctgctgttgaagcagcagcagc 25
      ||||| ||||| ||||| |||||
Db      289 CTGCTGTTGAAGTGCAGCAGCAG 269

```

```

RESULT 37
LOCUS   BH073174      601 bp      DNA      GSS      18-JUL-2001
DEFINITION
RPC1-24-315B1.TV RPC1-24 Mus musculus genomic clone RPC1-24-315B1,
DNA sequence.
ACCESSION BH073174
VERSION   BH073174.1 GI:14892771
KEYWORDS GSS.
SOURCE   house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 601)
AUTHORS   Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
TITLE     Mouse BAC End Sequences from Library RPC1-24
JOURNAL   Unpublished (1999)
COMMENT   Other GSSs: RPC1-24-315B1.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@igf.org

```

```

Clones are derived from the mouse BAC library RPC1-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 315 row: B column: 1
Seq primer: T7
Class: BAC ends.

```

```

FEATURES
source
Location/Qualifiers
1..601
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-24-315B1"
/clone_lib="RPC1-24"
/sex="Male"
/cell_type="Spleen/Braint"
/notes="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPC1-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

```

```

BASE COUNT      171 a      126 c      132 g      172 t
ORIGIN

```

```

Query Match      71.2%; Score 17.8; DB 13; Length 601;
Best Local Similarity 90.5%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY      5 ctgctgttgaagcagcagcagc 25
      ||||| ||||| ||||| |||||
Db      591 CTGCTGTTGAAAGTGCAGCAGCAG 571

```

RESULT 38
 BG642085/c 617 bp mRNA EST 23-APR-2001
 LOCUS
 DEFINITION
 BG642085 617 bp mRNA EST 23-APR-2001
 pgl1c.pk001.c6 Chicken liver cDNA library Gallus gallus cDNA clone
 pgl1c.pk001.c6 5' similar to g16912582 ref1NF_036524.11 peflin
 [Homo sapiens] dbj|BA84922.11 (AB018357) ABP32 [Homo sapiens]
 dbj|BA85163.11 (AB026628) peflin [Homo sapiens] dbj|BA91680.11
 (AK001420) unnamed protein product [Homo sapiens]
 gb|AAH02773.1|AAH02773, mRNA sequence.
 BG642085
 BG642085.1 GI:13774006
 EST.
 chicken.
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 1 (bases 1 to 617)
 Morgan, R.W., Burnside, J. and Cogburn, L.A.
 Chicken liver ESTs (2001)
 Unpublished (2001)
 CONTACT: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburn@udel.edu, www.chickest.udel.edu.
 Location/Qualifiers
 1..617
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /clone="pgl1c.pk001.c6"
 /clone_lib="Chicken liver cDNA library"
 /sex="Male and Female"
 /tissue_type="liver"
 /lab_host="E.coli EMD10B"
 /note="Vector: pCMVSPORT 6"
 /not="Vector: pCMVSPORT 6"
 BASE COUNT 111 a 205 c 196 g 99 t 6 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 617;
 Best Local Similarity 90.5%; Pred. No. 9.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctcgtgtttgaagcgcagcag 25
 ||||| ||||| ||||| |||||
 Db 406 CTGCTGATGAGGCCGACGACG 386

RESULT 39
 BG711664/c 646 bp mRNA EST 08-MAY-2001
 LOCUS
 DEFINITION
 BG711664 646 bp mRNA EST 08-MAY-2001
 pgl1n.pk008.m4 Normalized liver library Gallus gallus cDNA clone
 pgl1n.pk008.m4 5' similar to g16912582 ref1NF_036524.11 peflin
 [Homo sapiens] dbj|BA84922.11 (AB018357) ABP32 [Homo sapiens]
 dbj|BA85163.11 (AB026628) peflin [Homo sapiens] dbj|BA91680.11
 (AK001420) unnamed protein product [Homo sapiens]c, mRNA sequence.
 BG711664
 BG711664.1 GI:14005614
 EST.
 chicken.
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 1 (bases 1 to 646)
 Burnside, J., Morgan, R.W. and Cogburn, L.A.
 Chicken ESTs from a normalized liver library
 Unpublished (2001)
 CONTACT: Joan Burnside

FEATURES
 source
 1..646
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /clone="pgl1n.pk008.m4"
 /clone_lib="Normalized Liver Library"
 /sex="Male and Female"
 /tissue_type="liver"
 /lab_host="E.coli EMD10B"
 /note="Vector: pCMVSPORT 6"
 /not="Vector: pCMVSPORT 6"
 BASE COUNT 115 a 215 c 202 g 106 t 8 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 646;
 Best Local Similarity 90.5%; Pred. No. 9.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctcgtgtttgaagcgcagcag 25
 ||||| ||||| ||||| |||||
 Db 385 CTGCTGATGAGGCCGACGACG 365

RESULT 40
 BE912977 748 bp mRNA EST 29-SEP-2000
 LOCUS
 DEFINITION
 BE912977 748 bp mRNA EST 29-SEP-2000
 601668948F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3968890 5',
 mRNA sequence.
 BE912977
 BE912977.1 GI:10410127
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 748)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgap@femail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LIAW145 row: P column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..748
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3968890"
 /clone_lib="NCI CGAP Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="3 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 147 a 150 c 201 g 250 t

Query Match 71.2%; Score 17.8; DB 11; Length 748;
 Best Local Similarity 90.5%; Pred. No. 1e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ctgcctttgaagcgcagcag 25
 |||||||||||
 Db 567 CTGCTGTTTGAACACACGCG 587

RESULT 41

AM253374 172 bp mRNA EST 17-DEC-1999
 LOCUS
 DEFINITION UI-R-BJ0-aen-b-09-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone

ACCESSION AM253374
 VERSION AM253374.1 GI:6596965
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 172)
 AUTHORS Ronaldo M.F., Lennon G. and Soares M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msosares@blue.weeg.uiowa.edu
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-NO.

FEATURES

Source

Location/Qualifiers
 1..172
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BJ0-aen-b-09-0-UI"
 /clone_id="UI-R-BJ0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p7713D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0
 library is a subtracted library derived from the UI-R-AA1,
 UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
 UI-R-AG1 libraries. These libraries represent tissues from
 rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
 at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
 canal at 15 dpc, and ventricle at 13 dpc. The tag is a
 string of 5-6 nucleotides present between the Not I site
 and the oligo-dT track. The library was constructed as
 described by Ronaldo, Lennon and Soares, Genome Research
 6: 791-806, 1996.
 TAG_L1B=UI-R-BJ0
 TAG_TISSUE=ventricle at 15 dpc
 TAG_SEQ=GTGTC"
 BASE COUNT 30 a 51 c 38 g 53 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 172;
 Best Local Similarity 83.3%; Pred. No. 1e+03;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgcctttgaagcgcagcag 25
 |||||||||||
 Db 53 GATCTGCTTTCATGCCACGCTG 76

RESULT 42

AM121471 194 bp mRNA EST 22-OCT-1999
 LOCUS
 DEFINITION UI-M-BH2.2-aon-g-01-0-UI.s1 NIH_BMAP_M.S3.2 Mus musculus cDNA clone

ACCESSION AM121471
 VERSION AM121471.1 GI:6096804
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 194)
 AUTHORS Ronaldo M.F., Lennon G. and Soares M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H

National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
 available by the means that is soon to be determined. When NIH
 determines the means for distribution of the BMAP cDNA clones, this
 record will be updated accordingly when that means is determined.
 The following repetitive elements were found in this cDNA sequence:
 14-87, >GC-richlow-complexity
 Seq primer: M13 Forward
 POLYA-NO.

FEATURES

Source

Location/Qualifiers
 1..194
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH2.2-aon-g-01-0-UI"
 /clone_id="NIH_BMAP_M.S3.2"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p7713D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The
 NIH_BMAP_M.S3.2 library is a subtracted library of
 series, ultimately derived from a mixture of individually
 tagged normalized libraries from ten regions of the mouse
 brain (cerebellum, brain stems, olfactory bulbs,
 hypothalamus, cortex, amygdala, basal ganglia, pineal
 gland, striatum, hippocampus) after a series of
 subtractions to reduce the representation of cDNAs from
 which ESTs had already been generated. The following
 serially subtracted libraries were generated in this
 process: NIH_BMAP_M.S3.2, NIH_BMAP_M.S2, NIH_BMAP_M.S1.
 The subtracted library (NIH_BMAP_M.S3.2) was constructed
 as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2
 clones from which 3' ESTs had been derived was used as a
 driver in a hybridization with the NIH_BMAP_M.S2 library
 in the form of single-stranded circles. The remaining
 single-stranded circles (subtracted library) was purified
 by hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the
 NIH_BMAP_M.S3.2 library. This procedure has been
 previously described (Ronaldo, Lennon and Soares, Genome

TITLE
JOURNAL
COMMENT

T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamataka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resesc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermocyclization of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
Source

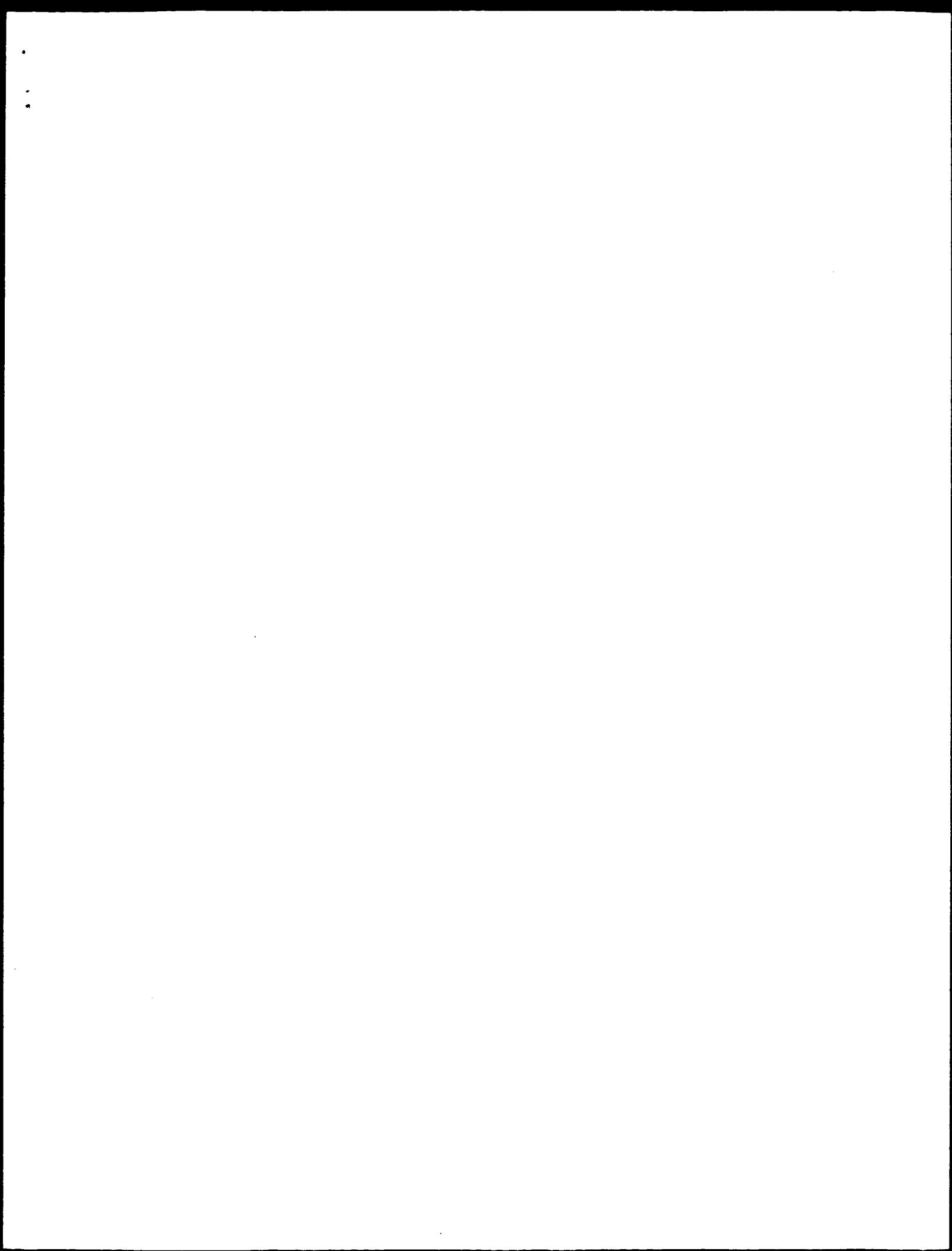
Location/Qualifiers
1. 310
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_11b="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCCGACACGAGCTTTTCTTTTCTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGACGTTATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 88 a 91 c 62 g 69 t
ORIGIN

Query Match 70.48; Score 17.6; DB 10; Length 310;
Best Local Similarity 83.38; Pred. No. 1.le+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcagcaag 25
||||| ||||||| |||||||
Db 77 gatcttcactttgaagccacag 54

Search completed: December 26, 2001, 12:11:32
Job time: 6454 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:56:35 : Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 ggcgcagcagtgatcgcgcagcagcalt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_ba:*
16: gb_fun:*
17: gb_hum:*
18: gb_in:*
19: gb_om:*
20: gb_ov:*
21: gb_pat:*
22: gb_ph:*
23: gb_pl:*
24: gb_pr:*
25: gb_ro:*
26: gb_sts:*
27: gb_sy:*
28: gb_un:*
29: gb_vl:*
30: gb_ba:*
31: gb_fun:*
32: gb_hum:*
33: gb_in:*
34: gb_om:*
35: gb_ov:*
36: gb_pat:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	AI1530	AI1530 B10b gene o
4	25	100.0	5793	1	ECOBIO	J04423 E.coli 7.8-
5	25	100.0	5872	6	A38246	A38246 Sequence 1
6	25	100.0	5872	6	A38251	A38251 Sequence 6
7	25	100.0	5872	6	A93674	A93674 Sequence 1
8	25	100.0	5872	6	A93679	A93679 Sequence 6
9	25	100.0	5872	6	ARI01809	ARI01809 Sequence
10	25	100.0	5872	6	ARI01810	ARI01810 Sequence
11	25	100.0	11022	1	AE000180	AE000180 Escherich
12	25	100.0	13501	1	AE005258	AE005258 Escherich
13	25	100.0	297816	1	AP002553	AP002553 Escherich
14	23.4	93.6	1121	6	E00893	E00893 Genomic DNA
15	23.4	93.6	5526	1	AF250776	AF250776 Unculture
16	20.4	81.6	12211	1	AE004895	AE004895 Pseudomon
17	20.2	80.8	3053	8	RIC0SAL	D10207 Oryza sativ
18	20.2	80.8	122052	8	AC079633	AC079633 Genomic S
19	20.2	80.0	2923	1	MAV250020	MAV250020 Mycobacte
20	19.2	76.8	12891	1	AE004192	AE004192 Vibrio ch
21	19.2	76.8	79419	9	HSJ324017	HSJ324017 Human DNA
22	19.2	76.8	349116	1	AP003003	AP003003 Mesorhizo
23	19	76.0	3603	1	PSEHRPIA	L11582 Pseudomonas
24	19	76.0	14846	1	AE007050	AE007050 Mycobacte
25	19	76.0	44201	1	MTCV180	297193 Mycobacteri
26	18.8	75.2	45704	1	AE001827	AE001827 Deinococc
27	18.8	75.2	110000	2	AL359456_2	Continuation (3 of
28	18.8	75.2	148900	2	AC025404	AC025404 Homo sapi
29	18.8	75.2	176687	2	AL365213	AL365213 Homo sapi
30	18.6	74.4	965	8	AF250770	AF250770 Unculture
31	18.6	74.4	1359	8	AF289025	AF289025 Cucumis s
32	18.6	74.4	3162	8	SRIDPHAI	X76536 S. tuberosum
33	18.6	74.4	3175	8	AF029256	AF029256 Kosteletz
34	18.6	74.4	3224	8	TOBPMAS3A	M80490 Nicotiana p
35	18.6	74.4	3229	8	TOBPMAS3B	M60166 Tomato (L.e
36	18.6	74.4	3233	8	TOBPMAS3C	M60491 Nicotiana p
37	18.6	74.4	3244	8	MCUBA891	AF275745 Lycopersi
38	18.6	74.4	3426	8	AF179442	AF179442 Lycopersi
39	18.6	74.4	13032	8	AF179442	AF179442 Lycopersi
40	18.6	74.4	14379	2	AE008283	AE008283 Agrobacte
41	18.6	74.4	33204	2	AC017995	AC017995 Drosophi
42	18.6	74.4	88548	8	ATT27H5	AL163852 Arabidops
43	18.6	74.4	103787	8	ATT27H5	AL138646 Arabidops
44	18.6	74.4	108879	8	ATT12K4	AC014784 Drosophi
45	18.6	74.4	120181	2	AC014784	AC014784 Drosophi
46	18.6	74.4	184278	3	AC0088221	AC0088220 Drosophi
47	18.6	74.4	186549	3	AE003777	AE003777 Drosophi
48	18.6	74.4	232000	3	AE003442	AE003442 Drosophi
49	18.6	74.4	306267	3	AE004631	AE004631 Pseudomon
50	18.4	73.6	10864	1	AE004959	AE004959 Pseudomon
51	18.4	73.6	10871	9	AC048343	AC048343 Homo sapi
52	18.4	73.6	33590	9	DME277193	AJ277193 Drosophi
53	18.2	72.8	1203	3	DME277175	X68347 Drosophila
54	18.2	72.8	1345	3	SZENIA	AJ277174 Drosophi
55	18.2	72.8	1486	3	DME277174	AJ277177 Drosophi
56	18.2	72.8	1671	3	DME277177	AJ277181 Drosophi
57	18.2	72.8	1725	3	DME277181	AK027288 Homo sapi
58	18.2	72.8	1729	3	AK027288	X75498 D. melanog
59	18.2	72.8	1746	9	DMRMAPA	AK024758 Homo sapi
60	18.2	72.8	1757	9	AK024758	X57435 H. sapiens m
61	18.2	72.8	1828	9	HSTFAP	AJ277179 Drosophi
62	18.2	72.8	1887	3	DME277187	AJ277187 Drosophi
63	18.2	72.8	1889	3	DME277187	AJ277188 Drosophi
64	18.2	72.8	1893	3	BC010576	BC010576 Homo sapi
65	18.2	72.8	2038	9	DME277186	AJ277186 Drosophi
66	18.2	72.8	2054	3	DME277186	AJ277182 Drosophi
67	18.2	72.8	2054	3	DME277182	AJ277176 Drosophi
68	18.2	72.8	2109	3	DME277182	
69	18.2	72.8	2109	3	DME277182	
70	18.2	72.8	2122	3	DME277176	

71 18.2 72.8 2141 3 DME277190 A1277190 Drosophila
72 18.2 72.8 2142 3 DME277191 A1277191 Drosophila
73 18.2 72.8 2149 9 S73885 S73885 AP-4-basic
74 18.2 72.8 2204 3 DME277192 A1277192 Drosophila
75 18.2 72.8 2209 3 DME277194 A1277194 Drosophila
76 18.2 72.8 2225 3 DME277183 A1277183 Drosophila
77 18.2 72.8 2330 3 DME277189 A1277189 Drosophila
78 18.2 72.8 2341 3 DME277185 A1277185 Drosophila
79 18.2 72.8 2530 9 BC009253 BC009253 Homo sapi
80 18.2 72.8 2547 9 AK002060 AK002060 Homo sapi
81 18.2 72.8 3113 8 SS95BASAL A19223 Thernus aqu
82 18.2 72.8 7954 1 TAO19223 AEO01900 Deinococc
83 18.2 72.8 10915 1 AEO04574 AEO04574 Pseudomon
84 18.2 72.8 12144 1 AEO04574 AEO04574 Pseudomon
85 18.2 72.8 18943 2 AC020236 AC020236 Drosophila
86 18.2 72.8 21420 1 SCF85 A111470 Streptomy
87 18.2 72.8 29625 1 SCF85 A1159178 Streptomy
88 18.2 72.8 35856 1 SCF85 A113851 Streptomy
89 18.2 72.8 40419 9 AC004653 AC004653 Homo sapi
90 18.2 72.8 59816 1 AF33038 AF33038 Streptomy
91 18.2 72.8 62198 2 AC013832 AC013832 Streptomy
92 18.2 72.8 72284 2 AC020328 AC020328 Drosophila
93 18.2 72.8 80389 3 AC005643 AC005643 Drosophila
94 18.2 72.8 84246 3 AC002512 AC002512 Drosophila
95 18.2 72.8 93615 2 AC004950 AC004950 Homo sapi
96 18.2 72.8 95064 2 AL592548 AL592548 Homo sapi
97 18.2 72.8 110000 2 LMFCHR36.09 Continuation (10 o
98 18.2 72.8 110000 2 LMFCHR36.30 Continuation (31 o
99 18.2 72.8 110000 3 AEO01572_2 Continuation (3 of
100 18.2 72.8 110000 3 AEO01572_3 Continuation (4 of

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew,
TITLE Enhanced biotin diosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source 1. 1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatgcagcatt 25
|||||
Db 79 GCGCAGCAGTGCATGCCAGCATT 103

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source 1. 1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatgcagcatt 25
|||||
Db 79 GCGCAGCAGTGCATGCCAGCATT 103

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL
FEATURES
source Patent: GB 2216530-A 16 11-OCT-1989;
Location/Qualifiers
1. 1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24. 1064
/gene="B10B"
24. 1064
/gene="B10B"
/transl_start=1
/codon_start=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-Prot:P12996"
/translation="MAHRPRWTLQVTELEFEKPLDLLFEAOVHROHPPROVOYST
LISIKTGACPEDEKCYCPQSSRYKGLAEERLMEVQVLSARAKAAGSTRPCMAW
KNPHERMPYLEOMVOGVKAMGLEACMTIGTSSQAOPLAANGIDYVHNINDTSEF
YGNLITPTRYOERLDPLFEKVRDAGTCVCSGIVGLAETVKDRAGLLQLANLPPTPES
VPINMLVKVKGCTPLADNDVDATDFIRITIAVARIMPSTVSRASAREQMNQOTAMC
PMAGANSITFGCKLITTPNEDEKDLQLEFRKGLNPOQTVLAGNDEQORLEOALMT
PUDVEYNNAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatgcagcatt 25
|||||
Db 102 GCGCAGCAGTGCATGCCAGCATT 126

RESULT 4
LOCUS ECOB10 5793 bp DNA
DEFINITION E.coli 7,8-diamino-pelargonic acid (bioa), biotin synthetase

(biob), 7-keto-8-amino-pelargonic acid synthetase (biop), bioc protein, and dehydrobiotin synthetase (biob), complete cds.
 J04423.1 GI:145422
 7.8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioc gene; biob gene;
 bioc gene; biob gene; biop gene; biotin synthetase; dehydrobiotin synthetase.
 Escherichia coli (strain K-12) DNA.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5793)
 Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.
 The Escherichia coli biotin biosynthetic enzyme sequences predicted
 J. Biol. Chem. 263, 19577-19585 (1988)
 8906784
 Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
 Location/Qualifiers
 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLRGDKLPHRYVNGMGYDGNISPHLAMDVPACIK SEVAVCYPDATPGSGMMHVVVNLPAITRVLPQFGSGLVAMPDGVQITRTDCKTS VGGAAPPGETHRVYTHALDIERIDVDEGASAMVFNHFHLSVASITAPMS"
 complement(633..1925)
 /gene="bioc"
 complement(633..1925)
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA23514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAFDQRIHHPYNSMTSPVPVPSAEGCELLSGRR LVDGSSWMAAHTGYNHPOLNAAKSODTAMSHVFGGTTTAPALELCKIKIYAMSGRN ALECVFLADSGVAVEVAMKALQYWAQGEARQRLTRNGTGGTGAMSCVCPDN SMHSLKGYLPENLEFAPQSRMDGEMDRDVGFRALMAARHEIAAVITPIVQGA GGRMYHPMLKRIKICDREGILLADELATGFGTKLFAECHEIAVLPDILGKA LVCMTLSATLTREVAETISNGEAGCFMHGTFPGNPLACAAASIALLESQMO QGVADIEVOLRDLAPARDAENVADVRVIGALGVETTPVMAALQKFFVEQGWIR PGKLIYMPPIILPQOLRLTAANRAVODEFTFCQ"
 2012..3052
 /gene="biob"
 2012..3052
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRPRWTLTSQVTELEFEPDLDLFEAQOVHRQHFDPQVQVST LLSIKTACPEDCRCPQTSRYKTGLEAEMLBEVULESARAKAAGSIRFCMGAM KNHERMPRLDEMOVCYKAMGLACMTIGTISSEAORLANAGLIDYNNINLPTSEF YGNTITRTYQERLDITLTKYKRGKIKYCSGIVGLSEYVYKDRAGLLLOLANIPTPES VPIMLKRVKGTPLADNDVDADFIRTAIVARIMPTSTVRLSAGREQNBETOQAMC FMAQANSIFYGCKLLTTPNDEEDKDLQLEFKGLINPQUTAVLAGDNEQOORLEQALMT PDDEYNNAAAL"
 3049..4203
 /gene="biop"
 3049..4203
 /gene="biop"
 /codon_start=1

/transl_table=11
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSWDEKINAALDARRADALRRPYPAVQAGGRHIVADROYLNF SSNDYLGSHHPOILIRAWQOAEQFGISGGSGHVSQVHQALEEELALMGTISRA LLEISGFANQAVIAAMAKEDRIADRSASHSLSEASLSQLRARRAHNDVTHLAR LLAPECGOONVYTEGEVSMDDSDAPLAEIQOVTQOIHGMIMVDADHGTGIGEGRG SCMLQKVPKELLVYTFKGEVSGAVALCSSVADYLIQFAHLLIYSTMPAQOAL RSLAVITRDEGDARREKIALITRRRAGVODLPPTLADSCSAIOLPLVGDMSRALQL AEKLRQGGWVTAIRPPYPAQTARLRLTLTAHEKODIDRLLEVLHONG"
 4190..4945
 /gene="bioc"
 4190..4945
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="MATVNRKQALYAAFGRAAHYEQHADLQROSDALLAMLPRKYT HVDAGCGPQMSRHMRERHAYVALDLSPPMLVQRQADADHYLAGDIESLPLATA TFDLAMSNLAVQWCGNLSALRELTVYVRPKGVAFPTTLVQSGSLERHQAOVDERP HANRFLPPDEITEOSLNGVYHOHIOTPTLMPDADLSAMSLKIGATFHLHEGRDPRIL TRSOLQRLQALAMPQOQGRITPLTHLFLVYAR"
 4938..5597
 /gene="biob"
 4938..5597
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="dehydrobiotin synthetase"
 /protein_id="AA23518.1"
 /db_xref="GI:145428"
 /translation="MSKRYPTGTDEEVGKTVASCALLQAAKRGYRTAGYKPVASGS EKPEPQDMVSLALQALONSSLODYAVVNYTFAEPTSPHIIAOGEGRTESLYMSAG LRAEQADRNLYVGVCGMPTLSDPTTFADWYTOBOLPVTLVGVKGLGINAMTLA QVIOHAGITLTAAGVANDYTPPKRHADYMTTLRMIDPRQWERSPLQKIQKKRQES T"
 1363 a 1554 c 1631 g 1245 t
 BASE COUNT
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.
 Query Match 100.0%; Score 25; DB 1; Length 5793;
 Best Local Similarity 100.0%; Pred. No. 1.4; Mismatches 0; Gaps 0;
 Matches 25; Conservative 0; Mismatches 0; Gaps 0;
 QY 1 ggcgcagcagtgcatgcgcagcatt 25
 DB 2090 gccgcagcagtgcatgcgcagcatt 2114
 RESULT 5
 A38246 5872 bp DNA PAT 05-MAR-1997
 LOCUS Sequence 1 from Patent WO9408023.
 DEFINITION A38246
 ACCESSION A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5872)
 Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 TITLE Patent: WO 9408023-A 1 14-APR-1994;
 JOURNAL LONZA AG (CH)
 COMMENT
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913

Other publication FI 951547 950331
 Other publication JP 85016947 960227.
 Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"
 1..96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23..28
 /evidence=experimental
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105..119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHRPRWTLSQVTELEFEKPLDLLPEAQVHQRHQPDPVQVST
 LLSKTGACPEDEKCPQSSRYRTGLEAEELMEVEVLESARAKAAGSTRFCMAAW
 KNPHERMPYLEQWVQVQKAMGLACMTLGTSQAOQLANGILLYNNIDTSEF
 YGNITITRYQERIDLTLEKVDAGIKVCSGIVGLGFTVRDAGILLQIANPTPES
 VPIIMLKVKGTPLADNDVDVDFPIRTIATVARIIMPTSVYRLSAGREONETQMC
 FMAGANSIFYGCKLITTPNEEDKDLQDLFRKLLNFOQTAVLAGNEQQRLEQALMT
 PDDEYNNAAAL"
 2284..2297
 /standard_name="BIOC RBS"
 2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVTKQALIAAFGRAAHYEQHADLIQOSADALLMLPORKYT
 HYLDACGCGWMSRRHREHAYTALDISPMLVQAROKDADHLAGDIESLPLATA
 TDLAMSNLAVQKGNLSTALKELETVVPRKGVAVFTLVQSLPELHQAWQAVERP
 HANRELPEDEISLNGVHGHIOPTILMPFDALSAMRSKLGIGATHLHEGRDRIL
 TTSQLORLQALMPQOGRYPLTYHLFLGVIARE"
 3742..3752
 /standard_name="BIOA RBS"
 3750..5039
 /gene="BIOA"
 3750..5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANATE AMINOTRANSF.""
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDLIAFDQHIHTHTSMTSPLPVYVVASBCEHLLSDGRK
 LVDGSSMMAIHGYNHPOLNAKRSQJIDAMSHVFGGITPAPELQRLVAMTPDP
 LECVFLADSGVAEVAEMKMLQYQWAKGEARQRLTFRNGYHGDTEGAMSVCPDPS

MSLAKGYLLENLAPAPSRMDGEDEBDWGFARLMAAHRHEIAVYIETVOCAG
 GMRVHPENIKRTRKICDREGLILLADELATGGRGLFCEHAEIADILCLAKAL
 TGCTMTLSATLITREVAEITSNCEACGFHGGTGMGNPLACAAANASIALIESGMOQ
 QVADIVQUREQJLAPARADMAVADVRLGATGVETHTPVNMAALQKFFVEGVIRP
 FKLILYIMPPYIILPOOLORLITAAVRAVODETFPCQ"
 5088..5100
 /standard_name="ORF1 RBS"
 5098..5574
 /gene="ORF1"
 5098..5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGLKPLHREHVGMDGDDNISPLAMPDPAKTK
 SEVVTGYPDPAFTGSGMWHMVYVNLPAITRVLPQFGSGIVAMPDVLTQTTDQKTKG
 YDGAAPKGETHTRTFTYHALDIRIDYDEGASGAMGCVNHFSLASASTVAMFS"
 5583..5605
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"
 1552 C 1695 G 1307 T

BASE COUNT 1318 A 1552 C 1695 G 1307 T

ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcgcagcagtcagtcgcagcatt 25
 |||
 Db 195 GCCGACGAGTGCATGCCGACGATT 219

RESULT 6
 LOCUS A38251 5872 bp DNA
 DEFINITION Sequence 6 from Patent WO9408023.
 ACCESSION A38251
 VERSION A38251.1 GI:2294849
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 6 14-Apr-1994;
 LONZA AG (CH)
 COMMENT
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 85016947 960227.
 Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A15-9"
 1141..1156
 /standard_name="BIOF RBS"
 1154..2308
 /gene="BIOF"


```

CDS
1154..2308
/feature="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence="experimental"
/product="KAPA SYNTHASE"
/protein_id="CA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKCTALNADARRADALRRRYPVACAGRWLVADROYLNF
SSNDYLGSHHPQITIRAMQCAEQFGSGSGHYSGYSVYHOALFEELAEWGYSA
LFIISGFANAOAVIAAMAKEDRTIADRLSHASLLEASISPSQDRRAHDVYHLAR
LLASPCPGQWVTEGVGSMDSAPLAIEQVTOQHNGWMLVDAGTGVIGEOGRG
SCWLOKXPELVITFGFSGVSAALVCSSTVADYLLQFAHRLIYSTMPAQAOL
RASLAVIRSDGDAAREKLAALITRFAGVODLPPTLADSCSAIQIPIVGNSTALOL
AEKLRQGCWYAIRPTVPYAGTARLRLTLTAHEMODIDRLLEVLHONG"

RBS
3030..3045
/standard_name="B10D RBS"
3043..3753
/feature="B10D"
3043..3753
/feature="B10D"
/gene="B10D"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence="experimental"
/product="DTB SYNTHASE"
/protein_id="CA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFTGTDTVEGKTVASCALLQAKAAGRYTAGIKPVASGS
EKTPGLRNSDALLORNSLQLDYATVNPYFAPTSPIISAOGRPIESLVNSAG
LRALQOAWMIVLEGAGGWFPTLSPFTFADWTOEOLPVLLVGVKIGCIHHMILTA
QYIHOAGLTLAGVAVANDVTPPGKRHAHEVMTLITMLPAPLGEIPLAENPENATGK
YINLAPVDASTIGFTSRL"

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0% Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggcgcagcagtgatcgccagcalt 25
      1 |||||
Db      195 gccgcagcagtgatcgccagcatt 219

RESULT 7
LOCUS      A93674      5872 bp      DNA
DEFINITION      Sequence 1 from Patent EP0798384.
ACCESSION      A93674
VERSION      A93674.1 GI:6741862
KEYWORDS
SOURCE
ORGANISM      Escherichia coli.
      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
      Escherichia
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL
      LONZA AG (CH)
FEATURES
Source
      1..5872
      /organism="Escherichia coli"
      /strain="DSM498"
      /db_xref="taxon:562"
      /clone="PBO30A-15/9"

promoter
1..96
/function="PROMOTER PTAC"
/evidence="experimental"
23..28
45..49
/standard_name="PROMOTER PTAC"
/evidence="experimental"
105..119
/standard_name="B10B RBS NO.9"
/evidence="experimental"
117..1157
/gene="B10B"
117..1157
/gene="B10B"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRRPMTLSQYTELEFEKPLDLLEFAQGVHROHEDPROVOYST
LLSKTQACEPDCKCYCQSSRYKTLAEERLEMEVQVLESARAKAAGSTRCMGAMW
KNPHERDMPYLBQWQVQKAMGLEACMTICTTSESOAORLANAGLDYVHNHNDTSPF
VGINITRTYQERLDPTLEKVDAGIKVCSGCTVGLGPTYKDRAGLLLOLANIPTPES
VPINMLVKVQTEPLADNDVDADFIRTLAVARIMPTSTVRLSAGREQNEQTOAMC
FMAGANSIFYGCKILITTPNPEDEKDIQIFRKLIINPOQAVVLADGNEQOORLEQALIT
PDPEYVNAAL"
2284..2297
/standard_name="B10C RBS"
2295..3050
/gene="B10C"
2295..3050
/gene="B10C"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATYVKQAIAAFGRAAAHYEQHADLQROSADALLAMLPOKRYT
HYLDGCGPGMSRHHREHROVATLDSPPMLVOAKOKDAADHYLAGDIESLPLATA
TFDLANSLAYQWQCNSTALRELXRVKPVAVAPTLVQGLPELHOAWOAVDERP
HANRRLPDELEQSLNGVYHOTHITPTTFDDALSAMRSIKGATLHHEGRPRL
TRSQLQRLQIAMPOQGRYPLTHLFGVIAKE"
3742..3752
/standard_name="B10A RBS"
3750..5039
/gene="B10A"
3750..5039
/gene="B10A"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLADFDRHIWHPYTSMTSPVYVYVVSABCECLILSDGR
LYDGSMMKATIHGYNHNPOLNAAKMSQIDAMSHWFGGITAPAPALECRKLVAMPPOP
LECVPLADSGSAVEVAMKMALQYQAKGEARQRLTPRNCYHGDTFCAMSCVDPNDS
MHLKMGYLPENLFAPOSRMDGEMDERDVGFKLAAHREHIAVITTEPIYOGAC
GMRYHPEWLKRIKRICDEGILLIADETATGEGRTGKLFACIEHAIAPDILCLKAL
TGGTMTLSATLTTRVAAETISNGEAGCEPHGPTFGNPLACAAVASLATIESGMOO
QVADIEVQLEOALPARDAENYADVAVIGALICVETTPYVNAALQKFEVGVWIRP
FKLLIYIMPEYIILPQQLRLTAAVNRAVODETFCCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/gene="ORFI"

```

```
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRGDKLPHRHVINGMCGYDCDNI SPLAMDDVPAIRK
SEVATCPDPAPTGSGMHWVYVNLPAOTRVLPGFSGSLVAMPDGLQETPEGKTG
VDGAAPRGETHRITFTVHALDIERIDVDEGASGAMGFVHFHSLASASITAFMS"
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgcagcagtcgcatcgccagcatt 25
Db 195 gccgcagcagtcgcatcgccagcatt 219

RESULT 8
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
141..1156
/standard_name="BIOT RBS"
1154..2308
/gene="BIOT"
1154..2308
/gene="BIOT"
/standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMOEKINAIADARRADARRRYPVAGAGRWLVADROYLNF
SSNDYLGISHRQITIRAWOGAEQFGISGSGSHVSGSVVHQALESLELAEMLGYSLR
LLEISGFANQAVIAAMAKEDRIADRLSHASLEASLSPSQLRFRANDVTHLAR
LSPSPGOQNVVEGVFSGMDSDAPLAEIQOQOHNHGMVLVDAHGTCVITGGGRC
SCMLQKVPPELLVTFGKGFGVGAAGVCSGVADVILLOFARHI IYSTMPRAQOAI
RASLAVIRSDGADAREKLAALITFRRAVCYODLFTPTLADSCSAIOLPIVGNSTALO.
AEKLRQCGCWTAIRPTVPAGTATRLTLTAHEMODIDRLLEVLHNG"
3030..3045
/standard_name="BIOT RBS"
3043..3753
gene
```

```
/gene="BIOT"
3043..3753
/gene="BIOT"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVGTDEGKTVASCALLQAAKAAGYRTAGYKPVASGS
KRPBELRNSDALAQRNSSLQIDYATVPYFAPRPSPHIISAQGRPISLVASAG
LRALDQADWVVEGAGWETPLSDFTFTFADWVTOEOLPVILVGVKIGCTNHMLTA
QVIOHAGILTLAGWANDVTPPEKRIAREYMTTLTRIRAPRLGEIPWLAENFENATGR
YINAEVDSATIGFTSRLL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgcagcagtcgcatcgccagcatt 25
Db 195 gccgcagcagtcgcatcgccagcatt 219

RESULT 9
ARI01809 5872 bp DNA PAT 14-FEB-2001
LOCUS ARI01809
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION ARI01809
VERSION ARI01809.1 GI:12812607
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
source
location/Qualifiers
1..5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgcagcagtcgcatcgccagcatt 25
Db 195 gccgcagcagtcgcatcgccagcatt 219

RESULT 10
ARI01810 5872 bp DNA PAT 14-FEB-2001
LOCUS ARI01810
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION ARI01810
VERSION ARI01810.1 GI:12812608
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
```

FEATURES
source
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gccgcagcagtgcatcgcgcagcatt 25
Db 195 gccgcagcagtgcatcgcgcagcatt 219
|||||

RESULT 11
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503

REFERENCE
AUTHORS 2 (bases 1 to 11022)
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
AUTHORS 3 (bases 1 to 11022)
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
AUTHORS 4 (bases 1 to 11022)
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HC01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@mar.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site

FEATURES
source
repeat_region
gene
CDS
protein_bind
gene
CDS

(<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

location/Qualifiers
1. 11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1. 130
/note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"
complement(147. 1430)
/gene="ybhC"
/note="b0772"
complement(147. 1430)
/gene="ybhC"
/function="orf; Not classified"
/note="1427; 98 pct identical to fragment YBHC_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"
/codon_start=1
/transl_table=11
/product="putative peptidesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVSLALALAFGVLTACSSNPDPQSDQTAAPTSSRPI LSAKKAONPDADHYFASITLPGAAANPSPITLPACDPFVCPAGTQGVHTTIOAVD AMRIKTNKROYIAVMPGEYGVYVPAAPAGGITLYGGEKPIDYKIGSLDGKSPA DWRHDVNPGRKIMPGKPAKPMYMDSCGSKSDRGVYCSAVFSSONNGLOLMTLENT LGDSVDAGNPAAVALRTDQVGINVNTLGRONTFFYVNSGVQARLETNRQPLVLT NSYIEGDVDIVSGRGAVFEDNTEFRVNSRTQOEAIVFAPATLSNITYGFLVNSRFN AFGDGVAAQLGRSLVDANTNGQVYIRDSAINEGFNTAKPMADAVISNRPAGTGSVD DNDIQRNLNDTVNRMMEYNNRGVSKVAEAKK"
complement(1507. 1535)
/note="factor Sigma70; predicted +1 start at 806574"
complement(1582. 2058)
/gene="ybhB"
/note="b0773"
complement(1582. 2058)
/gene="ybhB"
/function="orf; Unknown"
/note="1158; 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf; hypotheetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDGLKPLRHVFNCGMGYDGNISPLIADVDVPACTG SVFVTCYDPDPAFGSGMHWVYVNLPAQDVTYLVQGRSGIVAMPDGLQTRDQGTG YDGAAPKRGELHRIFTVHALDIERIDVDGASGAVGVHVRHSLASATIMPS"
complement(2108. 2124)
/note="central position to predicted promoter:85.5"
/bound_moiety="Rhas predicted site"
complement(2117. 3406)
/gene="bioA"
/note="b0774"
complement(2117. 3406)
/gene="bioA"
/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers: Biotin"
/note="142; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"

```

/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTDDIAFDORHIMPTSMSPLPVVSABCEILISDGKRL
LYDSSSMMAIHGNIHPQLNAAKSQIDAMSHWFGITHAPAEICRKLVAMPPOP
LECVLADSGVAEVAEAKMALQYWAQGEAKORLIFRNGHCDITGAMSCDPDMS
MISLTKGYPENLFAVAPQSRMDGEWDERDVGAFRLAARHETIAVILIEPVQAG
GMRMYHPEMLKRIKICREGILLIADIEATPGRTGKFLACAAASLALIESCDMK
TGGWTLSATLTREVAETISNGEAGCGMPTMGNPLPACAAASLALIESCDMK
QVADIEVQLEOLAPARDAEMVADVAVIGAVIETHPVMAALQKFFPEQGVIMRP
FGKILMPYIIIPQOLORTAANVAVQDETFPCQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"
/bound_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
/bound_moiety="BioB documented site"
3413..3441
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
/gene="4533"
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="0346; 99 pct identical to BIOB_ECOLI SW: P12996"
/transl_table=1
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPRYLSQVTELEFEKPLDLLFEAOVHROHPPROVST
LSIKTACPEDCKYCPOSSRYKTLGAERLMEVQVLSARAKAAGSTTRCGAAV
KNHERDMPEVLEOMVOGVKAMGLEACMTGLTSEGAORLANAGIDYVHNLDTPSE
YGNITRTYOEERDLTEKVRDAGIKVCGSGVIGCEYKDRAGLLDLQANLPTPEP
VPINMLVVKYKTPLEANDVDVADFIRTLAVARIMPTSVRLSAGREOMQOTAMC
PMAAGANSIFYGCKLITTPNEEDKDIQIFRKLGLNPQOTAVIACGNEQOORLEQALMT
PIDEYNAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="0384; 100 pct identical to BIOF_ECOLI SW: P12998"
/transl_table=1
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSMOEKIAIALDARRADALRRRYVAGAGRNILVADPROLYN
SNDYLGSHHPOILRAMOGKAEQFCGSGSGSVYSHQALEBELLAWLISRA
LLFISPAANOAVIAAMAKEDRIADRLSHASLSHSPQLRRFANDVYHLAK
ILASPCGOQWVTEGVSMQDSAPLAEIQOQVQHNGMLVADAHGTGVIGEGRGK
SCWLOKRPPELLVTEFGKFGVSGAAVLCSSVTADYLLQFARHLIYSTSMPPAQOAL

```

```

Query Match      100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3571 GCGCAGCAGGCGATCGCCGACGATT 3595

RESULT 12
AE005258
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED
11206551

TITLE
JOURNAL
MEDLINE
PUBMED
11206551

REFERENCE
AUTHORS
2 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-Oct-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
<1..7576
/note="O-island #36: Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomK"
/note="20981"
66..665
/gene="lomK"
/function="putative membrane; other or unknown (phage or
prophage related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
g317532789[gb]AAAF63331.1[AF151091_2 (AF151091)] lom
[prophage P-EB1a]"
/transl_table=1
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AA055137.1"
/db_xref="GI:12513752"
/translation="MRKVCAILISAICLAAGVPAMASEHOSTLSAGYLHASTDAPG
SDDLNGINVKYREFDTLGLITFSYANAEDQKTHVSDTRMHEDYVRNRFVYVAG

```

```

gene
CDS
    /product="unknown protein encoded by prophage CP-933K"
    /protein_id="AAG55141.1"
    /db_xref="GI:12513756"
    /translation="MKIPSLDSNPNFSAKAGYAPAPAPNAENYADYLDIGKRILP
    SAADLSNYESVIRAVHDSRSRLIDQHTVDIGTAVLVDLDSRSOTPEPDAYSIGHNK
    VHIGCIKRYNEELNEESSVKIDDIQSLTCELELYVGQEPITPEACEENDEEY
    VSFSAPTDSYEMPSQWQGLIHEIHHVTGSDPSQSGNLELPTGLARRAQAELG
    MSVPDFKGYAPEREAREHLRLRNALROAARHEENKAFPERLGTISDRYEASPDPT
    EYSAISNTGYCFIQOHDFPGLAINDNIDANQIDLYGAPYIFPFQVDKHNQ"
    5350..6231
    /gene="20989"
    /note="20989"
    5350..6231
    /gene="20989"
    /function="orf; Other or unknown (Phage or Prophage
    Related)"
    /note="No significant matches"
    /codon_start=1
    /transl_table=11
    /product="unknown protein encoded by prophage CP-933K"
    /protein_id="AAG55142.1"
    /db_xref="GI:12513757"
    /translation="MLSPYSVLGCSMNSLTPNRLTSPDNRVLSYRDAVHSDGAQY
    KGNKRIYRVATDNKFCYVRESHCPTNLLHRLGMPKGETSRKLEVLNAPSAM
    EGVVHNSRDLPPYDIAPPELPSVDYRLSPGVIGKGNAYVEDAPDQVTKM
    EFTTSQNEVTSERCFNOYVGASAEKTYGNNGDIIGIRMKINGESLINSIPRO
    ABHAIYDMEDRLKOKGILFVDTTETNVLDRAKNEFNPIDISSYVNSDRSWSQIMQ
    SYHGKODLISVLSKI"
    6462..7160
    /gene="20990"
    /note="20990"
    6462..7160
    /gene="20990"
    /function="orf; Other or unknown (Phage or Prophage
    Related)"
    /note="No significant matches"
    /codon_start=1
    /transl_table=11
    /product="unknown protein encoded by prophage CP-933K"
    /protein_id="AAG55143.1"
    /db_xref="GI:12513758"
    /translation="MRPTSLNVLHQSRSRSMSTDLIESLVKASSVONIKNNPOLRF
    OGTDHNTIYOIEALIGSTETGYVLNAISISIKLSETVVIHLNLSRGLVMARD
    IDAENHRTGSDPHCNALNAVEYCGGISTVVDHATIVHELIHVFHNHNGRLVES
    SRASQKYSLLLEFARYGLGAFSEVLESENKFHEITGMPTSTVXPDSALHDDNT
    VSLGFQDRLHPL"
    complement(7668..8144)
    /gene="ybhB"
    /note="ybhB"
    complement(7668..8144)
    /gene="ybhB"
    /note="ybhB"
    complement(7668..8144)
    /function="orf; Unknown function"
    /note="Residues 1 to 158 of 158 are 99.36 pct identical to
    residues 1 to 158 of 158 from Escherichia coli K-12 Strain
    MG1655; B07773"
    /codon_start=1
    /transl_table=11
    /product="orf; hypothetical protein"
    /protein_id="AAG55144.1"
    /db_xref="GI:12513759"
    /translation="MKLISDLRQDKLPHRHVFNMGNGYDGNISPHLANDVPAGTK
    SFVTCYDPDAPTSQGMHVVNVLPAIDTFLVPGGSGVLAAPDGLQTRIDFGTGTG
    YDGAAPKGETHRYIFVHALDERIDVDYGASGAMVGVHHSIASASITAFMS"
    complement(8203..9492)
    /gene="ybhA"
    /note="ybhA"
    complement(8203..9492)
    /gene="20985"
    /note="20985"
    2489..3469
    /gene="20985"
    /function="orf; Other or unknown (Phage or Prophage
    Related)"
    /note="Residues 13 to 326 of 326 are 55.73 pct identical
    to residues 18 to 331 of 336 from Genpept 118 :
    g145854371gb1A025465.11(A0125520_60 (A0125520)
    g145854371gb1A025465.11(A0125520_59 (A0125520)
    predicted by Glimmer [Salmonella typhimurium LT2]"
    /codon_start=1
    /transl_table=11
    /product="unknown protein encoded by prophage CP-933K"
    /protein_id="AAG55140.1"
    /db_xref="GI:12513755"
    /translation="MLSPRTTFFHNSVNIYVSSPCQTVSPAGKEVELKIDKPTILF
    OFEPNPERYKDEVPYIYNTKQHPILYDNTNARIESRPMKIFVDFGQSVNOKTAFS
    KLERFENVMIYREDVDFSNTRKLSLIDYHITCEQRLRTEDKDELINLEIKELR
    ELSKADSLISWAKRNHAMPDFRNALILKAGIIPKTYNTKNGISFEGGCIYLD
    MDMLILGKIGTIYAPDGLSMHVDNRDVSNIENSAITVRSNHPALLEGLSFMHSKVD
    AHRYDGLCKGVKKYFNFTPLHNYNHCFDIEFNHPNIIIMNTSQYTCSSW"
    3530..4522
    /gene="20986"
    /note="20986"
    3530..4522
    /gene="20986"
    /function="orf; Other or unknown (Phage or Prophage
    Related)"
    /note="No significant matches"
    /codon_start=1
    /transl_table=11

```

```

Query Match
Best Local Similarity 100.0%; Score 25; DB 1; Length 13501;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ggcagcaggtgcatccgcagcatt 25
|||||
DB 9657 GCGACAGCAGTCATCGCAGCATT 9681

```



```

source
1. .5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/notes="unknown organism, cosmid clone derived from
environmental consortium"
complement(52. .528)
/notes="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kda protein in modc-bioa
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLTSLNDLRDGLPHRRVFNQMGVDCDNI SPLHAWDVPAATK
SFVATCYDPAATGSGMHWVYVNLPAATRVLPQSGSLVAMPDGVLOTETDFGKAG
YDGAAPKGETHRYLFVHALDIERIDVDEGASGAMGVNHPHSLASASTITAFS"
complement(587. .1876)
/gene="bioa"
complement(587. .1876)
/gene="bioa"
/notes="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase bioa"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFDGRHIMHPYTSMTSPVPYVVSAGECELLSDGRK
LVDMSSWMAAIGHYNHPOLANAKRSO IDAMSHVMEFGITHAPAIELCRKLVTPTD
KLCVFLADSGVAEAVAMKALQYQAKGROEFLFRNGRHGDFRNGSVCCDNG
MHSIMKGLPENLPAPAPQSMDEMERMGVAFRLMAHRRHTAVITTEFVQVCG
GMRVHPEMLKRIKICDRGILLIADETLNGRGTCKLACEAEIAPIDICGAL
TGGIMTSLATITTPREVAETTSNGACFPNHPGRTNGRLACAAANSLATESGDWQ
OVADIEVQLRQALAPADAENVADRVLCAGVETTHPVNMAALQKFEVEQGVWTRP
FGKITLMPPTITLPQQLRLTAHVNAVDETFECQ"
1963. .3003
/gene="bioB"
1963. .3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase bioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRRPMTLSQVTELEPERLDLTFEAOVHRRHEDPRVOYST
LTSIKGACPEDECKYCPQSSRYKTLGAERLMEVQLSARAKAKAGSTRCMGAW
KNENRDMPELEQMOGVKDLGLACMTCTLSGSAQRLANGIDYNNHNDTSPER
VGNITTRTYOEFLDLEKVRDAGIVCSGIVGLDETVKDRAGILLDLANLPTPEF
VPLNMLVKYKGTPLANDVDADFIRITAVARIMPTSYVRLSAGREOMEDQAMC
FMAANSIFYGCKLITTPNEFDKDLQFRKIGLNPQTAVALAGDQOORLEQALMT
PDTDEYYNAAL"
3000. .4154
/gene="bioF"
3000. .4154
/gene="bioF"
/notes="8-aminoc-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase bioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWOKTINAAIDARRADALRRRYPVAGAGRWLVADROYLINI
SSNDYLGUSHPOITIRAWOOGAEFGTIGSGSGSVSVHQALEELAEIAGLYSAR
LPIISGFANQAVIAMMAKEDRIADRLSHASLSLEASLSPQRRFPAHNDVTILAR
LIASPCGQOLVTEGVFSMDGSAPLAETIQOVTOOHNMWMLVDDAGHTVGEQGRG
SCMLQKPELVTEFGKGVSGAAVLCSSVADYLLQPAHILITSTSMPEAOQAL
RASLAVIRDEGDDARRKELVSLIAHRAVGVODLPTLADSCAIOPLIVGDNSTRQDL
AKLRQCGCWVAIRPPYPACTARLRLTITTAHMODIRLLEVLHENG"
4141. .4896
/gene="bioc"
4141. .4896
/gene="bioc"
/gene="bioc"

```

```

FEATURES
FEATURES
BASE COUNT      1274 a      1507 c      1567 g      1178 t
ORIGIN
Query Match
Best Local Similarity 93.6%; Score 23.4; DB 1; Length 5526;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ggcgaagagtgcatgcgcagcatt 25
Db 2041 GCGCAGCAGGTCATGCGCAGCATT 2065
RESULT 16
AE004895
LOCUS      AE004895      12211 bp      DNA      BCT      30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 456 of 529 of the complete
genome.
ACCESSION  AE004895 AE004091
VERSION    AE004895.1 GI:9951083
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 12211)
Stover,C.C., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z., and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 12211)
Stover,C.C., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Retzer,J., Sater,M.H.,
Hancock,R.E.W., Jory,S., and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers

```


[illegible]

Query Match	81.6%;	Score 20.4;	DB 1;	length 12211;
Best Local Similarity	95.5%;	Pred. No. 1.3e+02;		
Matches 21; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

RESULT	17
LOCUS	RICOSA1/c
DEFINITION	RICOSA sativa OSAL mRNA for H-ATPase, complete cds.
ACCESSION	D10207
VERSION	D10207.1 GI:218178
KEYWORDS	H-ATPase; OSAL: hydrolytic enzyme.
SOURCE	Oryza sativa (strain:Nipponbare K-1) callus cDNA to mRNA.
ORGANISM	Oryza sativa

BASE COUNT	807 a	649 c	779 g	818 t
ORIGIN				

	Query Match	80.8%;	Score 20.2;	DB 8;	Length 3053;
	Best Local Similarity	88.0%;	Pred. No. 1,8e+02;		
	Matches 22;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
Oy	1	gcgcagcaggtgcattccacgacatt	25		
Db	412	GCGCAGCAGCAGCATTGCCAGCATT	388		

RESULT	18	
LOCUS	AC079633/c	
DEFINITION	AC079633	122052 bp DNA
ACCESSION	AC079633	Genomic Sequence for <i>Oryza sativa</i> , Nipponbare strain, clone
VERSION	AC079633.9	OSJNBa0032208, from Chromosome 3, complete sequence.
KEYWORDS	AC079633.9	GI:13899390
SOURCE		
ORGANISM	<i>Oryza sativa</i> .	
PROJECT		01-MAY-2001

FEATURES	Location/Qualifiers
source	1. .122052
	/organism="Oryza sativa"
	/strain="Nipponbare"


```

/organism="Vibrio cholerae"
/strain="N16961"
/db_xref="taxon:666"
/chromosome="1"
/plotype="E1.10r"
/complement(84, .1370)
/gene="VC1111"
/complement(84, .1370)
/gene="VC1111"
/notes="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:U00096; identified by sequence similarity;
putative"
/transl_table=1
/product="adenosylmethionine-8-amino-7-oxononanoate
amino transferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMDLAFDRCHIMHPTSTLPLTCYPSRADGVMITLFGRSI
VDGSSMAAIGHYNNPKLNOAHQIDMSHTMFGGITHOPALICOKLILKAPNL
OOPLADSGSAVEYSLSKMAIQWMSKQBPRAKTLTLHGTHGDTFAMSVTDDNSM
HSLKGLPEHIFANSPEGGFDMADERDIADEPRHKLTEHHQIAAVILEIYVAGS
MRYHPEFLROYRALCDEGVLLILDEIATGFGTKLFCENHAGIQPDILCYOKALT
GGYMTLSATLTTOQVADIVCAGACGEMHGPFGMGNPLACVASASLSLEOGMOCQ
VANIEAFPAEOLPKLNDSPRYKOTRWLGAIGVETHLPVNMEVIQALFVHGVITRF
GRLIYMPYISOPHEHLOKLVDMAMAALNPDCKQ"
1497, .2549
/gene="VC1112"
1497, .2549
/notes="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEVRHNWTVAEVAKALLDPRMDLFEACQVHRLHHPHNOVST
LSTKTCACPEDCKYCPQSAHYRTDVDRKMEVEFRLDAOKKNSGSRFCGAAM
KNPKERDRLKEMIRGVKDKGLCTCMTLGMITLTDQOOLAQGLDYNNELTSPFE
YGNITTTTODRLDLSHVADGAKICSGGITIMKPSAVRKLISAGREKNEQOALC
VPIIMLVKVTGLEQVDVDFPDEKLIIVARIMKPSAVRKLISAGREKNEQOALC
FMAGANSIFVCGCKLITTPNPAEDSDMLFKKLCINRQVAKPDEITENEILDVVER
VAAPRTASDLFYDAAL"
2533, .3687
/gene="VC1113"
2533, .3687
/notes="similar to GB:J04423 SP:P12996 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRFKSRLEOALNERESHSLTRLOLDRSAOGLLIQODRTYL
NFSNDYLGLANDAEIVQWOTGLARFGAGSGASPVYTFSSAAHELEHLGMLGFE
RAVLSSGFSANQALLFTLLEKEDLLQDRIINAHSMACMLSPATMKRKNDTHL
ROLLHOSNSLYVEEGVSMDCAPLAQATILYKORAKMLVYDHAISIGVLSGDSGAS
COAGSIHPELIVTFGKAFGLAGAAVLCDNAHGDYLTQFARHHVYSTAMPADAHAT
HAARMTOSQWRROELAEILACDEQCRNIPGVATQTPILKPMILSSDSALASHEIL
KOQGIWVASIRPPIVPGSARLRITITAAHTQAIKRLSEQLKHMKEKP"
3549, .4487
/gene="VC1114"
3549, .4487
/notes="similar to GB:J04423 SP:P12996 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/organism="Vibrio cholerae"
/strain="N16961"
/db_xref="taxon:666"
/chromosome="1"
/plotype="E1.10r"
/complement(84, .1370)
/gene="VC1111"
/complement(84, .1370)
/gene="VC1111"
/notes="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:U00096; identified by sequence similarity;
putative"
/transl_table=1
/product="adenosylmethionine-8-amino-7-oxononanoate
amino transferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMDLAFDRCHIMHPTSTLPLTCYPSRADGVMITLFGRSI
VDGSSMAAIGHYNNPKLNOAHQIDMSHTMFGGITHOPALICOKLILKAPNL
OOPLADSGSAVEYSLSKMAIQWMSKQBPRAKTLTLHGTHGDTFAMSVTDDNSM
HSLKGLPEHIFANSPEGGFDMADERDIADEPRHKLTEHHQIAAVILEIYVAGS
MRYHPEFLROYRALCDEGVLLILDEIATGFGTKLFCENHAGIQPDILCYOKALT
GGYMTLSATLTTOQVADIVCAGACGEMHGPFGMGNPLACVASASLSLEOGMOCQ
VANIEAFPAEOLPKLNDSPRYKOTRWLGAIGVETHLPVNMEVIQALFVHGVITRF
GRLIYMPYISOPHEHLOKLVDMAMAALNPDCKQ"
1497, .2549
/gene="VC1112"
1497, .2549
/notes="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEVRHNWTVAEVAKALLDPRMDLFEACQVHRLHHPHNOVST
LSTKTCACPEDCKYCPQSAHYRTDVDRKMEVEFRLDAOKKNSGSRFCGAAM
KNPKERDRLKEMIRGVKDKGLCTCMTLGMITLTDQOOLAQGLDYNNELTSPFE
YGNITTTTODRLDLSHVADGAKICSGGITIMKPSAVRKLISAGREKNEQOALC
VPIIMLVKVTGLEQVDVDFPDEKLIIVARIMKPSAVRKLISAGREKNEQOALC
FMAGANSIFVCGCKLITTPNPAEDSDMLFKKLCINRQVAKPDEITENEILDVVER
VAAPRTASDLFYDAAL"
2533, .3687
/gene="VC1113"
2533, .3687
/notes="similar to GB:J04423 SP:P12996 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRFKSRLEOALNERESHSLTRLOLDRSAOGLLIQODRTYL
NFSNDYLGLANDAEIVQWOTGLARFGAGSGASPVYTFSSAAHELEHLGMLGFE
RAVLSSGFSANQALLFTLLEKEDLLQDRIINAHSMACMLSPATMKRKNDTHL
ROLLHOSNSLYVEEGVSMDCAPLAQATILYKORAKMLVYDHAISIGVLSGDSGAS
COAGSIHPELIVTFGKAFGLAGAAVLCDNAHGDYLTQFARHHVYSTAMPADAHAT
HAARMTOSQWRROELAEILACDEQCRNIPGVATQTPILKPMILSSDSALASHEIL
KOQGIWVASIRPPIVPGSARLRITITAAHTQAIKRLSEQLKHMKEKP"
3549, .4487
/gene="VC1114"
3549, .4487
/notes="similar to GB:J04423 SP:P12996 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1

```


/product="d3j24017.1.2 (A novel protein similar to
Drosophila CG11840, isoform2)"
/protein_id="CAC1947.1"
/db_xref="GI:11121544"
/translation="NADMPETITSRDARFPPIASCTLLGLYLFKIPSOEYINIL,
SMFVGLIALSHTISPENKFPSPFNROYOLFFOGSENKEEILFNPEDKIL
VCLGSSIVGVWYLKRHIAANFLGFLASGLNGVLLHNNRSCILLGGFTIDV
WVEGTINVTAKSFAPRIKIVFODLLEKGEANNFMLGIDVYIGITALLRP
DISLKNHTYFVTSFAAVIRGLGITITIMHFKHAOPALVLPACIGFVLAVALAK
GEVTEMFSTESNPPADPAVTESEKTEASASKLEKKEK"
1915..2083
repeat_region
/note="AluNb repeat: matches 134..300 of consensus"
2084..2374
/note="AluSg repeat: matches 1..291 of consensus"
2375..2503
/note="AluNb repeat: matches 5..134 of consensus"
2593..2788
/note="AluNb/FRAM repeat: matches 197..308 of consensus"
2835..3142
/note="AluNb repeat: matches 1..309 of consensus"
3232..3366
/note="AluSg repeat: matches 1..122 of consensus"
3367..3595
/note="AluSc repeat: matches 83..306 of consensus"
3596..3689
/note="AluSx repeat: matches 122..219 of consensus"
3696..3721
/note="Alu repeat: matches 90..115 of consensus"
3784..3826
/note="Alu repeat: matches 251..293 of consensus"
4002..4306
/note="AluSc repeat: matches 1..303 of consensus"
4439..4634
/note="AluSc repeat: matches 171..308 of consensus"
5360..5570
/note="AluNb repeat: matches 84..298 of consensus"
5743..6042
/note="AluSg repeat: matches 1..300 of consensus"
6791..6818
/note="MER88 repeat: matches 433..459 of consensus"
6819..7124
/note="AluSx repeat: matches 1..304 of consensus"
7125..7393
/note="MER88 repeat: matches 75..433 of consensus"
7444..7475
/note="MER88 repeat: matches 1..31 of consensus"
7476..7775
/note="AluNb repeat: matches 1..302 of consensus"
7776..8099
/note="MER88 repeat: matches 31..353 of consensus"
8100..8399
/note="AluNb repeat: matches 1..302 of consensus"
8402..8704
/note="AluY repeat: matches 1..302 of consensus"
8705..8789
/note="MER88 repeat: matches 353..447 of consensus"
8884..8959
/note="38 copies 2 mer aa 64% conserved"
9286..9467
/note="L2 repeat: matches 2559..2750 of consensus"
9679..9961
/note="AluSg repeat: matches 1..283 of consensus"
9985..10643
/note="L2 repeat: matches 1971..2748 of consensus"
10684..10999
/note="L2 repeat: matches 2146..2461 of consensus"
11951..12071

Query Match 76.8% Score 19.2; DB 9; Length 79419;
Best Local Similarity 87.5% Pred. NO.3.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgcatcgccagcat 24
Db 16785 GCCCAGCAGCGTGCAGCGCAGAAAT 16762
|||||
AP003003
LOCUS AP003003 349116 bp DNA BCT 15-MAY-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 10/21.
ACCESSION AP003003 BAO000012
VERSION AP003003.2 GI:14024067
KEYWORDS
SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.
ORGANISM Mesorhizobium loti
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
AUTHORS Watanabe,A., Ideasa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimo,S., Sugimoto,M.,
Takeuchi,C., Yamada,M. and Tabata,S.
TITLE Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
JOURNAL DNA Res. 7 (6), 331-338 (2000)
MEDLINE 21082930
REFERENCE 2 (bases 1 to 349116)
AUTHORS Kaneko,T.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
On May 11, 2001 this sequence version replaced gi:11994978.
COMMENT
FEATURES
Location/Qualifiers
1..349116
/organism="Mesorhizobium loti"
/strain="MAFF303099"
/db_xref="taxon:381"
/complement(54..1655)
/gene="mli13875"
complement(54..1655)
/gene="mli13875"
/codon_start=1
/transl_table=1
/product="phosphoglycerate dehydrogenase"
/protein_id="BAB50672.1"
/db_xref="GI:14024068"
/translation="MAPRVLVSDKLSPPFAVQFKDQGVEDYIPDLCKDKKLEKLEVID
OYDLATRSATKTYTEKILNATRIKIVGRGICVDVNDVDPAAASKIIVNTPFGNSI
TTAHAVAMFALAKRQIDFANASTHAGKWKPNFQVETIGKTLGIVGGNGISIVAT
RGVGLKMHVAFDPFLSDKRAEELGVEKVELDELAFARADFTLHTPLTKTRNITDAG

AIAAKNGVRLINCAKGLLVEADLIALKSKVAGAGIDVFEVEPAEONALGEMV
 VATPHLGASTAEQENVALQVEMSDYLIRKGAVSNAIMNPSTIAEEAPRIKPFVKLA
 EVLGAPGOVTEDEPIMEVELFDGSTATMTNTRALISATLGLAIRPOVSDVNMVAPIM
 VKREGIIVAEVKRKDKSGVDGYIKLIVKTEHRTSITAGTCFSDGKPRF10IKGINLDA
 EVGCHMLYTNADAPGIIIGLIGTCGNGNINANFQICRRNRPGSDAIALLYLDAPFRE
 KVLDOVRHAKSIDSAKRLQEPVGGI"
 complement(1719. .2894)
 /gene="m113876"
 /complement(1719. .2894)
 /gene="m113876"
 /codon_start=1
 /transl_table=11
 /product="phosphoserine aminotransferase"
 /protein_id="BAB50673.1"
 /db_xref="GI:14024069"
 /translation="MTHTKPKGLRPANPFSPPCAKPCGWSVEALNNALGRSHAK
 IGKTKLQALREILREILOVPADHRIGIVPASDTGAVEMALMSLIGERGVDVAMEFG
 SGWTVDYKOIKLPDVRKLEFAGCGELPDIAKIDPDROVYFTWNGTSGVVRNGDFIP
 ADKGLITICDTSAPFORIDPEKIDVYTESMOKVYLGEGAHGMLIISPRVARELESY
 KPAWLPKIFRLTSGGLIEGTFEGETIINTPMCLVEDYLDALHMAKSIIGLDLVAR
 ADANAAVLDGCVDKSSMLGLAVOPATRSNTSVCLPTDPVVALDADGCTFKAGIV
 SALDEGVADYIGSYRDAPGRLIMWCATVETSDLETLPLWDMAFAAOKASLAKAA"
 complement(3096. .3869)
 /gene="m113877"
 /complement(3096. .3869)
 /gene="m113877"
 /codon_start=1
 /note="hypothetical protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB50674.1"
 /db_xref="GI:14024070"
 /translation="MTLTSRYLALAGLLPLPAAVADYDPPIYADQADPVVPEV
 GSGWLRGDVSTLAKQSEKNDFAFTPASFDEKEDPITASIGFGYHFNLDLRADINLG
 YLPGNKIIGYGDYSLVVPATVAASDLKNYAVSLMNAVVDLGTVGYITPLVGGG
 VGIYOSTRLSANFTNNQDPTDVFQDDKTKYSALYLANGLAVOYKSNVSDLYG
 OYFSADEYVTAASLTSPVHKGISNQIKGLRYDLN"
 complement(4111. .4944)
 /gene="m113878"
 /complement(4111. .4944)
 /gene="m113878"
 /note="hypothetical protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB50675.1"
 /db_xref="GI:14024071"
 /translation="MPTARKALALLFAGLASPVFADLPPEPVEEAPPYEOPVD
 VCGWYIRCDIDYHKSVDRCIDYMTYCIDPCNCTVLPQKSFYGLKGFSLGGVGY
 KINDRLRADVTADYKFSNNGSTSDLYSTSEVMSALLILANAYVDIGTWGIIITP
 YVAGTGGARVKMDYDYPNTAETNPGASNMFPAAVAMGASCLTDKIILDAGYRFS
 HIOGGMFEMDASSAGPFGDGINTEHVEKGLRYOFGGNNNGCAAPVVAOPEPEPTIT
 K"
 complement(5193. .6545)
 /gene="m113879"
 /complement(5193. .6545)
 /gene="m113879"
 /codon_start=1
 /transl_table=11
 /product="phosphoglucosyltransferase/phosphomannomutase family
 protein: MraA"
 /protein_id="BAB50676.1"
 /db_xref="GI:14024072"
 /translation="MAGNFGTGDIGIRANKPMTAEIIMRVCMAGLSFQSGSHHR
 VLGKOTRLSGYMIENAMVAGIACAQMDVILCPITPAVAMIYRLRIDIGMISAS
 HNYVDNGIKLPEPDGKLSDETEERIESMLDDEVTLLADSGILRARKRVCHVRI
 IERAKRLPMSMSLGLIRIVDCANGAAYVPAEALMEGAEVVAIVINTEGNCNLIKE
 CGSTHPAGLQKVEHVAADIGIALDDADARVIVDENGAIVDDQIMAMIAESWHDSG
 RLAGGVYSTVMNGLERFLGDMKQLHRTKQGRVVEHMRHGLINMGCEGSGHIV
 LSDSTGDSVSLQVLACIKROGRVSELSKPEFVOLLKNVRJAGCKPLEEAPV
 KAATEDARNRLKAGRLVTRPGTEPLIRVMAAGDDPOLVEAVVNDIVERTSBSAA
 6913. .7779
 /gene="m113880"
 6913. .7779
 CDS

/gene="m113880"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB50677.1"
 /db_xref="GI:14024073"
 /translation="MLIHRLATLTGLAVTFLAAPANALTMKCESTRYNKADAGTL
 AGCTNMFOKACGTDAAATPTTPATTAETKPAKATETKPAKAAAAADSTAKGL
 TSKECSAKYQAAKAACTLNGMKMNDPRKTECAAGASAADPTKPAATTKPAATKAAAP
 AGCKGLIMAECSAKTQAAKTANTLKQCKKNNDRKAECCGAAAADDSDVAPSPANTKE
 PAKPTTVARKGVTPSAISPKYASLTPGKRMHTCLDQYIALKANALGGLWIKGG
 GFEYSLCNARKL"
 8079. .8537
 /gene="m113881"
 8079. .8537
 /gene="m113881"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB50678.1"
 /db_xref="GI:14024074"
 /translation="MPLIKLAVIATVASYAMSVAPASALTMKCESTRYNAKAGTLN
 GMDAAFRKKECAITPAESVSGDTVTYTTQPKKETDANAAVPAAPAGVFPPTIAAEF
 KTEKPAKARKKTCLOGYHGKKEAGTLNGLWIKGGGYISLCNARKABS"
 complement(8624. .10552)
 /gene="m113882"
 complement(8624. .10552)
 /gene="m113882"
 /codon_start=1
 /transl_table=11
 /product="metalloprotease (cell division protein): FlsH"
 /protein_id="BAB50679.1"
 /db_xref="GI:14024075"
 /translation="MNPYRNIALMAITIAVLLALENLFPOTPRGASDPVSOFLQ
 DVACGRKVTYIACARTITGYTNSNGSFQYSGDPLVRLDKNTIARREADGS
 NSLPGYLSIMPLILGVMTFPRQKQSSGRAMGSKAKULLTEHRTVPEOYA
 GVDAKEDLEIEFLRDPQFORLGGKIRGVLLVGPETGTLAKSVAGANVPE
 FTISGSDFVEMDFEVSASRVDMFDQAKNAPCIIFIDEIDAVGRHGAQLGGNDER
 EOTLNOLLIVEMDFEVSNEIILLIATNRPPVLLPALLRPFROVVVPPDLYGRK
 ILKVVHNVPLAPVNDLKVARCGTPGSGDLMMLVESALMAARKRLVTVAAEED
 AKKQIMGAERSSAMQAKELIYAHACHALLALNVPASADPLHKATIIIPRACLOM
 VMQLPBGDRYSMTKYMIISLATIMGRVNAEERFEGENTTSASSDIPQATILARM
 VTRWGFSDKLGVAVYGDQEEVFLGHSVARTONISETAQITIDAEVRLIDEXYSTRK
 SILTKKKKEWIALAOLGLEYLTSDEIKOLIAEKPARDLGDDTPPSKSAVPSGG
 RKKGPPEGCEGPEQPS"
 complement(10675. .12039)
 /gene="m113884"
 complement(10675. .12039)
 /gene="m113884"
 /note="Similar to yaeN protein [Bradyrhizobium japonicum],
 contains similarity to cell cycle protein MesJ/cytosine
 deaminase-related protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB50680.1"
 /db_xref="GI:14024076"
 /translation="MLDTEPDLSTRPLSFHIDFTGAAVAVSGSDSTALLLLKHNFD
 RTGSSAKLAVYIDHGLRQGSAAEQAAVAKCAEGRALHRTLTMTGRPSGLPAAAR
 DARVIAFLAARAEIGLIVGHTRDDADAEVYLMHMRHAGRELADLAGRAGAPATL
 DVMREYVIRPLUGTRRSALROFLREHNGMADPNIDEAFERRVRAALAGSAGNM
 ENTTLRACQAAVERGOLGASANLIRIRIASPSGTLILDPALLIADQAAIVYLRIL
 LATAGAAFLPDQARSALFVRLKGLFCATLSRVVDFRIGAGLREAREGPPPAT
 AVDNTIMGRRIITLNDSGALLIAPLGAARARLADIDGETPALSILRAALIAAEPTLL
 QAFENIGLPQCEPRLPVAARVYVSPFARPLSPDLAPARAVAEILGASPLPAPLGG
 HSAD"
 complement(12098. .13210)
 /gene="m113886"
 complement(12098. .13210)
 /gene="m113886"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB50681.1"
 /protein_id="BAB50681.1"

```

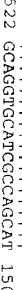
/db.xref="GI:14024077"
/translation="MHLRSVLSGTLAALLISGFPALASGFCASGPGOSTDGSFSEHJL
SLELPISLFGDKKKRPDOYVFAOOSNGAGATGLEDLRQWNRKILEELNFOQLLOMOEMRK
QOEENEFERFOOLEEGSAOOGOPPARPKSPATNTNTDYAAAPATAGAPADAGAAPGNSGS
AGGTGVEDVIVIESPEGSPGKVIIPGTGAPEKFGITVDKNNATVINDAGANTOSTPPOD
AAPATKAPAKAGKNGTIVAAIPSTNDEEELIYRNSYFOILISGDYGTACGGCRHSHSPH
PRDKAADAHYWLGESILLGQOKYRDAAEAVFLAASKDIPKAKKAPDMLLKGLGVSLVGLK

```

RESULT	23				
PSHRP1A/c					
LOCUS	PSHRP1A	3603 bp	DNA	BCT	21-OCT-1993
DEFINITION	Pseudomonas syringae hrp10 genes, complete cds.				
ACCESSION	L11582				
VERSION	L11582.1	GI:151267			
KEYWORDS	hrp1 gene; protein translocation.				
SOURCE	Pseudomonas syringae (strain 61) DNA.				
ORGANISM	Pseudomonas syringae				
	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.				
REFERENCE	1 (bases 1 to 3603)				
AUTHORS	Huang,H.-C., Xiao,Y., Lin,R.-H., Lu,Y., Hutcheson,S.W. and Collier,A.				
TITLE	Characterization of the Pseudomonas syringae pv. syringae 61 hrp1 and hrp1 genes: homology of Hrp1 to a Superfamily of proteins associated with protein translocation				
JOURNAL	Mol. Plant Microb Interact. 6, 515-520 (1993)				
FEATURES	location/Qualifiers				
Source	1..3603				

BASE COUNT	772 a	1093 c	1035 g	703 t
ORIGIN				

Query Match	76.0%;	Score 19;	DB 1;	Length 3603;
Best local Similarity	100.0%;	Pred. No. 5.9e+02;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
0y	6	gcagtgatcgcgcagcat	24	

Db	1522	GCAGGTGTCATCCGACCAT	1504
			
RESULT 24	AE007050/c	AE007050	14846 bp DNA
LOCUS	Mycobacterium tuberculosis CDC1551,	BCT	27-APR-2001
DEFINITION	complete genome.	section 136 of 280 of the	
ACCESSION	AE007050 AE000516		
VERSION	AE007050.1 GI:13881590		
FEATURES			

REFERENCE	1 (bases 1 to 14846)
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uitterlacker, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishal, W.
TITLE	Whole genome comparison of <i>Mycobacterium tuberculosis</i> clinical and laboratory strains
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 14846)
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uitterlacker, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishal, W.
TITLE	Direct Submission
JOURNAL	Submitted (23-Apr-2001)
FEATURES	The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Source	Location/Qualifiers 1. 14846

```

gene
145..1302
/gene="MT1935"
CDS
145..1302
/gene="MT1935"
/notc="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46208.1"
/db_xref="GI:13881591"
/translat="MRCELDLPVLGLSTPTTGLGVLAEHGSDGA ILDRNELTHASQ
RNMADHTEAOLAEVLLAEVAAAGDHRRLRVIGYMPNAPASQALLVETSTGAGF
NVNVPVRIATETLADALAPVIGYAOICAEHESATYVYNDTHGKQ IAVKRVG
GIGLSWTLMGMRGDNRPRAGVYVYSGDSVESESMQERLVPVPEQIMQAVKVC
RGALAAKASTETFDKQLYADSVSQPTVAPRSRHTAGAAALAAAVTAVASLSLA
GICLAEHNTGIAKAGAHPTPRILAKAALVAPVPPVLPVPAAPRPAAGHEPAA
VTSGELVTFEPNEEQPNAPASAPQDNRDNDQPTVTRLEHIFGAYGSAPPAE"
complement(1272..2237)
/gene="MT1936"
complement(1272..2237)
/notc="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46209.1"
/db_xref="GI:13881592"
/translat="MPRTWITAGNLRYPLSYNCQMPDMYOCRRHRPESCIOPARITAT
RSGVSAARRMTRPCACADEROMAWLVDRADAHKPSITQAPVYVVRGCDIPALSS
WQMLVWMLAIPIHYIVLEFLHVAWVYVYVIAFAAILPTGRYPTTLPDNGVMKRMK
VAFVYLSAGCDRPPPSLQTKAEYPADLLEVDYPERLSQGLVILKMWLLAIPIHYILTA
VELSSQMKRFLIDPHDRGVITMPPSLVILLVAVALLETGRVIGLVNVLIGNRNA

```



```

DEFINITION   Mycobacterium tuberculosis H37Rv complete genome; segment 85/162.
ACCESSION    Z97193 AL123456
VERSION       Z97193.1 GI:3261816
KEYWORDS
SOURCE        Mycobacterium tuberculosis H37Rv.
ORGANISM      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
               Actinomycetales; Corynebacteriinae; Mycobacteriaceae;
               Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE     1 (bases 1 to 44201)
AUTHORS       Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
               Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
               Tekle,A.F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
               Connor,R., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
               Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
               Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
               Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
               Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
               Barrell,B.G.
TITLE         Deciphering the biology of Mycobacterium tuberculosis from the
               complete genome sequence
JOURNAL       Nature 393 (6685), 537-544 (1998)
MEDLINE       98295967
REFERENCE     2 (bases 1 to 44201)
AUTHORS       Parkhill,J.
TITLE         Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
               tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
               Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de GenetiQue
               Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
               75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
               On Jun 27, 1998 this sequence version replaced gi:2225942.
COMMENT       Notes:
               Details of M. tuberculosis sequencing at the Sanger Centre are
               available on the World Wide Web.
               (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
               been renumbered from the original cosmid submissions but the old
               gene designations are in brackets after the new gene numbers.
               Gene prediction was based on a Hidden Markov Model of TB genes
               implemented in TBparse (Krogh) supplemented with visual inspection
               of positional base preference in codons, especially where there is
               an increase in the observed/expected third position G + C.
               CAUTION: In some cases we may not have predicted the correct
               initiation codon. Where possible we choose an initiation codon
               (arg, gtc, or ttg) which is preceded by an upstream ribosome
               binding site sequence (optimally 5-13bp before the initiation
               codon). If this cannot be identified we choose the most upstream
               initiation codon.
FEATURES             location/qualifiers
SOURCE               1..44201
                    /organism="Mycobacterium tuberculosis H37Rv"
                    /strain="H37Rv"
                    /db_xref="taxon:83332"
                    <1..42142
                    /organism="Mycobacterium tuberculosis H37Rv"
                    /strain="H37Rv"
                    /db_xref="taxon:83332"
                    /clone="Y180"
                    23..460
                    /gene="Rv1873"
                    23..460
                    /gene="Rv1873"
                    /note="Rv1873, (MTCY180.45c), len: 145. Unknown, TBparse
                    score is 0.909"
                    /transl_table=1
                    /product="hypothetical protein Rv1873"
                    /protein_id="CAB10052.1"
                    /db_xref="GI:2225987"
                    /db_xref="SPTREMBL:O07756"
                    /translation="MKSASDPPDLKRFVYAGAPVYRSVVEELRAGKRGHMMVEFEPD
                    LRKGSSPLAVRYGISLEADQAYLOHLDLGRLEHCTGLVNVQVGRSIEELFGPPDD
                    LKCSMTLEFARATDANQDFVALLAKYGGGGRKTVALLAVT"
RBS               517..520
                    /note="possible RBS, GGAG, for Rv1874"
                    533..1219
                    /gene="Rv1874"
                    533..1219
                    /gene="Rv1874"
                    /note="Rv1874, (MTCY180.44c), len: 228. Unknown, TBparse
                    score is 0.928"
                    /codon_start=1
                    /transl_table=1
                    /product="hypothetical protein Rv1874"
                    /protein_id="CAB10067.1"
                    /db_xref="GI:2225986"
                    /db_xref="SPTREMBL:O07755"
                    /translation="MIMREPPDDMCARQAVDALLGICVAGISTVNRSTVDSL
                    MTLTITPVPAVAVSILMPOCCGEQVAAALRLADCBELATVETSPLEFSLVE
                    SCGRRTGLANIALIRPPDLDDATVITLRQROHDTQVALDAQTFSTQVNVVVALTPE
                    ARGIACTVEELFPPVAATTDLKAFGAADNDLNRISRWVASTAFGANONIDVPTS
                    RYVFRTPFKD"
                    1218..1221
                    /note="possible RBS, GAGG, for Rv1875"
                    1230..1673
                    /gene="Rv1875"
                    1230..1673
                    /gene="Rv1875"
                    /note="Rv1875, (MTCY180.43c), len: 147. Some similarity
                    to29584[MTC165.22 (147 aa) opt: 178 z-score: 240.3 E):
                    7.4e-06] 26.9% identity in 130 aa overlap. TBparse
                    score is 0.908"
                    /codon_start=1
                    /transl_table=1
                    /product="hypothetical protein Rv1875"
                    /protein_id="CAB10051.1"
                    /db_xref="GI:2225985"
                    /db_xref="SPTREMBL:O07754"
                    /translation="MTTLNEAALAAARGLAVSTVADCTQASLVNGLPHRPS
                    GPSSLPFTYGVKVGKIGNLRAPQLAVTRNOMNTVGRQLGPPDPRLVDR
                    IRLRLRYTAAAGGTHDDMDERDMAEDQRAVVALTPRTLSNG"
                    2179..2183
                    /note="possible RBS, AGGAG, for Rv1876"
                    2189..2668
                    /gene="bfrA"
                    2189..2668
                    /gene="bfrA"
                    /note="Rv1876, (MTCY180.42c), bfrA. len: 159. Function:
                    bacterioferritin. FASTA results: BFR.MYCLE P43315
                    bacterioferritin (bfr) (159 aa) opt: 958; E(): 0; 90.6%
                    identity in 159 aa overlap. TBparse score is 0.913"
                    /codon_start=1
                    /transl_table=1
                    /product="bfrA"
                    /protein_id="CAB10050.1"
                    /db_xref="GI:2225984"
                    /db_xref="SWISS-PROT:O08465"
                    /translation="MGCDPDLRLNEDQITSELTAINOYFLHSMQDNMGFTELAHT
                    RAESPDREHAEITRILLIDGLPNQIRGSRIGOTLEQFPADATIDVLDNRK
                    PGIVMCREKQDTSAAVLLEKVADEEHIDLEQLDLMKIKIGELYSACVSRPP"
                    2742..2746
                    /note="possible RBS, GGAGG, for Rv1877"
                    2753..4816
                    /gene="Rv1877"
                    2753..4816
                    /gene="Rv1877"
                    /note="Rv1877, (MTCY180.41c), len: 687. Integral membrane
                    protein similar to many antibiotic and drug efflux
                    proteins. M. tuberculosis relatives include: MTCY3612.01,
                    MTCY98.0002c and MTCY369.27c. Contains P500217 sugar
                    transport proteins signature 2 (P500217). FASTA results:
                    056175 (557 aa) opt: 895; E(): 0; 34.7% identity in 528
                    aa overlap. TBparse score is 0.916"
                    /codon_start=1
                    /transl_table=1
                    /product="hypothetical protein Rv1877"

```

```

/protein_id="CAB10049.1"
/db_xref="GI:2225983"
/translation="MAGPAPPTAPTAIRAGCPHLSFRRNIIETALVFCVLAAGQ
TIVPALPTVAELGSDVDMASVTLGGTGVVVVAGLGLGLGRLVGLSVF
VSCVGLGSLQTMMLAISRALGVGASISVAVLAALVPLRDGRGVLGAVF
GVNTVTPPLGLGMLTDLNRMWFINVPSIAVLTVAATVAALAPKPPVLDYGI
LVIAVATLIMATSWGTGYAWGSATVGLIGAVAGLGFVWLEGRANAALIPRL
FGSPVAVCCVLSVVGFMALGALTPVILGVYDASATASGLRLLIPVIGLILAST
GTGVLVGRTRKIFPVAGMALMAVAFLLSOMDEMPPILQSLIVLIGAGLISQ
VLIVLQNTSSFDLGATSGVTFEFPVVGASFGATFGLFVNLRLGSLATGSAF
PVPAVSPVAILHOLPQSMAPVIAVAESLTOVFLCAVTVVGLFLLALREPLD
IHDAADLGGCFVPAESPEDYLEIAVRMLPGVLRDLAQPGCGGLAVELMAL
RITVYORLFEAVRLTDLGRHLHVPYQVEFEPFDLVDTGAARDGLITLTPSGHRV
DSLAVLIRWMLDLHAAVAPGLKRPDHPHQAOLHVDVAVLQVWEDJLSESKO
LAATP"
/misc_feature
3122..3199
/gene="Rv1877"
/feature="PS00217 Sugar transport proteins signature 2"
/gene="glnA3"
4871..6223
/gene="glnA3"
4871..6223
/gene="glnA3"
4871..6223
/feature="Rv1878. (MTCY180_40c), len: 450. glnA3. Similar to
many glutamine synthetases. FASTA results: GlnA_BACCE
P19064 (EC 6.3.1.2) (443 aa) opt: 497; E(): 5.2e-23; 29.0%
identity in 331 aa overlap. Tblast score is 0.925. Also
similar to C-terminus of Fluc_EMENT p38094 flug protein.
emerellaniidians (865 aa), fasta scores: opt: 227
z-score: 357.9 E(): 6.4e-13 (29.9% identity in 394 aa
overlap) (note that the downstream ORF MTCY180_39c is
similar to the N-terminus) and to several other M.
tuberculosis proteins eg. MTCY427.03c, MTCY427.01 and
MTCY190.31"
/codon_start=1
/product="glnA3"
/protein_id="CAB10048.1"
/db_xref="GI:2225982"
/feature="SPTREMBL:007752"
/translation="MTATPLAAATAIALEACVDTVIGTVVNPAGLTOAKTPIRPN
TEANGLGASPVWTFECIDGSIATFADISVGDORLRLDLSALRLIGDGLAMPAGF
FEODSTPVACSRGTLRLTAAADAGIDAVIGHEVEFLVADADGRLPSTLMQYQ
ACVLEHEAFVRVNAATAAGALAEOPHPEHAGANFETSIAPOPVAAADLVITRLI
IGTRARRHGLRVLSLPPAPAGISGAGHOFSLTMSSEGLSCGTGAAGMTSAGFAV
AGVLRGLPDAQOILGSIIVGLRMPPGMAWAGIYACWGNREAAVREKGGASAYGG
NVEKVVDPSPANPYLASAAILGLALDGKTKTAVLPSETTVDPQLSDVDRAGILRL
AADQADAIIVLIDSSKILRLCLGDPVAVAVVQLHERYGLDLPQALADKFRMAVSV
"
RBS
6213..6216
/gene="glnA3"
/feature="possible RBS, GGAC, for Rv1879"
6226..7362
gene
76.0%; Score 19; DA 1; Length 44201;
Best Local Similarity 100.0%; Pred. No. 4,6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 cgcacaggtgcacgcaca 20
|||||
Db 25769 CGCAGCAGTGCATCGCA 25751

RESULT 26
AE001827 45704 bp DNA circular BCT 10-JAN-2001
LOCUS AE001827/c
DEFINITION Deinococcus radiodurans R1 plasmid cpl, complete plasmid sequence.
ACCESSION AE001827
VERSION AE001827.1 GI:6460959
KEYWORDS
SOURCE Deinococcus radiodurans.
ORGANISM Deinococcus radiodurans
Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

REFERENCE
AUTHORS 1 (bases 1 to 45704)
White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,
Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,
Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J. J., Lam, P., McDonald, L., Uitterback, T., Zaleski, C.,
Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W.,
Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S.,
Smith, H. O., Venter, J. C. and Fraser, C. M.
JOURNAL Genome sequence of the radioresistant bacterium Deinococcus
MEDLINE radiodurans R1
PUBMED Science 286 (3444), 1571-1577 (1999)
10567266
2 (bases 1 to 45704)
White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,
Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,
Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J. J., Lam, P., McDonald, L., Uitterback, T., Zaleski, C.,
Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W.,
Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S.,
Smith, H. O., Venter, J. C. and Fraser, C. M.
JOURNAL Direct Submission
MEDLINE Submitted (03-NOV-1999) The Institute for Genomic Research, 9712
PUBMED Medical Center Dr, Rockville, MD 20850, USA
10567266
FEATURES
source
1..45704
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/plasmid="Cpl"
73..384
/gene="DRC0001"
73..384
/gene="DRC0001"
/feature="identified by sequence similarity; putative"
/codon_start=1
/product="cytochrome P450-related protein"
/translation="MCSPPRRYSGSDAETLLPVYAEPSPVYHIDISOANDRESVF
TQAOQCFPHKKNVRLHSLFGKGHACLGSLVTTTCRVMAHLELLSARAGORTDYS
0"
725..1951
/gene="DRC0002"
725..1951
/gene="DRC0002"
/feature="identified by sequence similarity; putative"
/codon_start=1
/product="phosphoenolpyruvate synthase-related protein"
/translation="MSHLVRLIEDISLTGIPDDELGRILQETIFELCGAYDHHRLTP
EVEAFYMWCRINLAGVTNEHEHFLQDFLSLRQAGFEKLVIAHAKSSITLGEVDEDS
DMAIQOPFLAEISQVSKHGILYITTYCDIGRPSEWEDPSAIVIAIKRYMANLPDLSL
RIFQOILKROBCOVLSAALPDIKRPEDRLYSASFSLHREDHVEMLQDRALALR
RVCKAFGRHLALGLIEDPLDCAVFTTELTYOACGISIAQALRELPARKSYESDLK
FKPRPTREAPRSSPDVSKLGMGTSGRGSVSGSVSGVSGAFKLSMDILVCP
ETGPDWTPURSINAGLYTESFAGGMSHALVAREYIRPAIILGKSKSLDRYYDL
DGTGLVETIGRFESE"
1938..2801
/gene="DRC0003"
1938..2801
/gene="DRC0003"
/feature="identified by sequence similarity; putative"
/codon_start=1
/product="phosphoenolpyruvate synthase-related protein"
/translation="MNCDDCKNSLCLDSCGIPSLPLFLVDRMLERVGKAGANLGE
TAGMFEIPNCVITIDVFAALNAVSLPVGGRKRYGSPFNHMFPODLVYHLKEMMER

```


Qy 4 cagcagcagtcagcagcagcalt 25
 ||||| ||||| ||||| ||||| |||||
 Db 22665 CAGCAGAACATCGCCAGCATT 22686

RESULT 28
 AC025404
 LOCUS AC025404 148900 bp DNA HTG 24-MAR-2000
 DEFINITION Homo sapiens chromosome 1 clone RP11-301H2 map 1, WORKING DRAFT
 SEQUENCE, 17 unordered pieces.
 AC025404
 AC025404.2 GI:7321650
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 148900)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 1, clone RP11-301H2
 Unpublished
 2 (bases 1 to 148900)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, C.,
 Campopiano, A., Castle, A., Choeppel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howard, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,
 Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehotzky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margolis, N.,
 McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McNeely, R.,
 Melrim, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Sudamanian, A., Talamas, J.,
 Testafie, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 24, 2000 this sequence version replaced g1:7210127.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: W1BR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L8110
 Center clone name: 301_H2

Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 139855 bases at least Q40
 Consensus quality: 144213 bases at least Q30
 Consensus quality: 145912 bases at least Q20
 Insert size: 151000; agarose-fp
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces

is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1610: contig of 1610 bp in length
 * 1611 1710: gap of 100 bp
 * 1711 4083: contig of 2373 bp in length
 * 4084 4183: gap of 100 bp
 * 4184 6794: contig of 2611 bp in length
 * 6795 6894: gap of 100 bp
 * 6895 10293: contig of 3399 bp in length
 * 10294 10393: gap of 100 bp
 * 10394 14279: contig of 3886 bp in length
 * 14280 14379: gap of 100 bp
 * 14380 17741: contig of 3362 bp in length
 * 17742 17841: gap of 100 bp
 * 17842 21194: contig of 3353 bp in length
 * 21195 21294: gap of 100 bp
 * 21295 27169: contig of 5875 bp in length
 * 27170 27269: gap of 100 bp
 * 27270 34236: contig of 6967 bp in length
 * 34237 34336: gap of 100 bp
 * 34337 42772: contig of 8436 bp in length
 * 42773 42872: gap of 100 bp
 * 42873 52351: contig of 9479 bp in length
 * 52352 52451: gap of 100 bp
 * 52452 63753: contig of 11302 bp in length
 * 63754 63853: gap of 100 bp
 * 63854 74258: contig of 10405 bp in length
 * 74259 74358: gap of 100 bp
 * 74359 87809: contig of 13451 bp in length
 * 87810 87909: gap of 100 bp
 * 87910 102652: contig of 14743 bp in length
 * 102653 102752: gap of 100 bp
 * 102753 119718: contig of 16966 bp in length
 * 119719 119818: gap of 100 bp
 * 119819 148900: contig of 29082 bp in length.

FEATURES
 source
 misc_feature
 /clone="RP11-301H2"
 /clone_lib="RP11-11 Human Male BAC"
 1..1610
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 1711..4083
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 4184..6794
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 6895..10293
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 10394..14279
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 14380..17741
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 17842..21194
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 21295..27169
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 27270..34236
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 34337..42772
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 42873..52351
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 52452..63753
 /note="assembly_fragment"
 misc_feature
 /note="assembly_fragment"
 63854..74258
 /note="assembly_fragment"

```

misc_feature          /note="assembly_fragment"
                       74359..87809
misc_feature          /note="assembly_fragment"
                       87910..102652
                       /note="assembly_fragment"
                       clone_end:SP6
                       vector_side:right"
misc_feature          /note="assembly_fragment"
                       102753..119718
                       clone_end:T7
                       vector_side:left"
misc_feature          /note="assembly_fragment"
                       119819..148900
BASE COUNT          43768 a 28628 c 29209 g 45693 t 1602 others
ORIGIN
Query Match          75.2%   Score 18.8; DB 2; Length 148900;
Best Local Similarity 90.9%   Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 4 caccagatgcacgcacgatt 25
      ||||| ||||| |||||
Db 74753 CAGCAGACATCCGACGCAATT 74774

RESULT 29
AL365213             176687 bp      DNA      HTG      23-JAN-2001
LOCUS               Homo sapiens chromosome 1 clone RP11-224019, *** SEQUENCING IN
DEFINITION          PROGRES ***; 18 unordered pieces.
ACCESSION           AL365213
VERSION             AL365213.10 GI:12227406
KEYWORDS            HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE              human.
ORGANISM            Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE           1 (bases 1 to 176687)
AUTHORS             Pavitt,R.
TITLE               Direct Submission
JOURNAL             Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
                    CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk
                    On Jan 14, 2001 this sequence version replaced gi:12225456.
COMMENT
                    Genome Center
                    Center: Sanger Centre
                    Center code: SC
                    Web site: http://www.sanger.ac.uk
                    Contact: humgeny@sanger.ac.uk
                    Project Information
                    Center project name: BA224019
                    ----- Summary Statistics -----
                    Assembly program: XGAP4; version 4.5
                    Sequencing vector: plasmid; 108752; 100% of reads
                    Chemistry: Dye-terminator Big Dye; 100% of reads
                    Consensus quality: 167635 bases at least Q40
                    Consensus quality: 170858 bases at least Q30
                    Consensus quality: 172773 bases at least Q20
                    Insert size: 174987; sum-of-contigs
                    Insert size: 160662; 29.8% error; agarose-fp
                    Quality coverage: 3.77x in Q20 bases; sum-of-contigs Quality
                    coverage: 4.31x in Q20 bases; agarose-fp
                    -----
                    * NOTE: This is a 'working draft' sequence. It currently
                    * consists of 18 contigs. The true order of the pieces
                    * is not known and their order in this sequence record is
                    * arbitrary. Gaps between the contigs are represented as
                    * runs of N, but the exact sizes of the gaps are unknown.
                    * This record will be updated with the finished sequence
                    * as soon as it is available and the accession number will
                    * be preserved.
                    * 1 19212: contig of 19212 bp in length
                    * 19213 19312: gap of 100 bp

```

*	19.13	36455:	contig of 17143 bp in length
*	36456	36555:	gap of 100 bp
*	36556	41590:	contig of 5035 bp in length
*	41591	41690:	gap of 100 bp
*	41691	47894:	contig of 6204 bp in length
*	47895	47994:	gap of 100 bp
*	47995	50477:	contig of 2483 bp in length
*	50478	50577:	gap of 100 bp
*	50578	57800:	contig of 7223 bp in length
*	57801	57900:	gap of 100 bp
*	57901	71528:	contig of 13628 bp in length
*	71529	71628:	gap of 100 bp
*	71629	76173:	contig of 4545 bp in length
*	76174	76273:	gap of 100 bp
*	76274	105540:	contig of 22657 bp in length
*	105541	105640:	gap of 100 bp
*	105641	111894:	contig of 6524 bp in length
*	111895	111994:	gap of 100 bp
*	111995	123775:	contig of 11781 bp in length
*	123776	123875:	gap of 100 bp
*	123876	132910:	contig of 9035 bp in length
*	132911	133010:	gap of 100 bp
*	133011	136379:	contig of 3369 bp in length
*	136380	136479:	gap of 100 bp
*	136480	144593:	gap of 8514 bp in length
*	144594	145093:	gap of 100 bp
*	145094	153649:	contig of 8556 bp in length
*	153650	153749:	gap of 100 bp
*	153750	157198:	contig of 3449 bp in length
*	157199	157298:	gap of 100 bp
*	157299	161364:	contig of 4066 bp in length
*	161365	161464:	gap of 100 bp
*	161465	176687:	contig of 15223 bp in length

RES	location/Qualifiers
source	1. 176687
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP11-224019"
	/clone_1lb="RPc1-11.1"
misc_feature	1. 19212
	/note="assembly fragment: 014656"
misc_feature	fragment_chain:1"
	19313. 36435
	/note="assembly-fragment: 011227"
misc_feature	fragment_chain:1"
	36556. 41530
	/note="assembly fragment: 001555"
misc_feature	fragment_chain:1"
	41691. 47894
	/note="assembly-fragment: 002227"
misc_feature	fragment_chain:1"
	47995. 50477
	/note="assembly-fragment: 001774"
misc_feature	fragment_chain:1"
	50578. 57800
	/note="assembly-fragment: 013411"
misc_feature	fragment_chain:1"
	57901. 71528
	/note="assembly-fragment: 004522"
misc_feature	fragment_chain:1"
	71629. 76173
	/note="assembly fragment: 01981"
misc_feature	fragment_chain:2"
	76274. 105540
	/note="assembly-fragment: 00319"
misc_feature	fragment_chain:2"
	105641. 111894
	/note="assembly-fragment: 01486"
misc_feature	fragment_chain:2"
	111995. 123775
	/note="assembly-fragment: 00713"
misc_feature	fragment_chain:3"

```

misc_feature 123876..132910
              /note="assembly_fragment:01542
              /fragment_chain:3"
misc_feature 133011..136379
              /note="assembly_fragment:00535"
misc_feature 136480..144993
              /note="assembly_fragment:00623"
misc_feature 145094..153649
              /note="assembly_fragment:01468"
misc_feature 153750..157198
              /note="assembly_fragment:01546"
misc_feature 157299..161364
              /note="assembly_fragment:02010"
misc_feature 161465..176687
              /note="assembly_fragment:00187
              /clone_end:17
              /vector_side:right"
BASE COUNT 52919 a 35491 c 34728 g 51832 t 1717 others
ORIGIN
Query Match 75.2%, Score 18.8; DB 2; Length 176687;
Best Local Similarity 90.9%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 cagcagcagtcgcgcgcgcatt 25
Db 58230 CAGCAGGACATCGCCAGCATT 58251

```

```

RESULT 30
LOCUS AF250770 965 bp DNA BCF 31-JAN-2001
DEFINITION Uncultured bacterium pCosHE1 DAPA-aminotransferase (bioc) and
            biotin synthase (biob) genes, partial cds.
ACCESSION AF250770
VERSION AF250770.1 GI:12620104
KEYWORDS
SOURCE uncultured bacterium pCosHE1.
          Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 965)
            Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streif,W.R.
            Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
            Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL MEDLINE
PUBMED 2 (bases 1 to 965)
11133432
Entcheva,P., Liebl,W. and Streif,W.R.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
JOURNAL Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
            Location/Qualifiers
FEATURES
            source
            1..965
            /organism="uncultured bacterium pCosHE1"
            /db_xref="taxon:143796"
            /clone="pCosHE1"
            /note="unknown organism, cosmid clone derived from
            environmental consortium"
            complement(<1..438)
            /gene="biob"
            complement(<1..438)
            /gene="bioc"
            /note="7,8-diaminopelargonic acid
            synthetase-aminotransferase; similar to bioc"
            /codon_start=1
            /transl_table=11
            /product="DAPA-aminotransferase"
            /protein_id="AAG60563.1"
            /db_xref="GI:12620105"

```

```

/translaction="MTQDDLAFDQOHIMHPYTSMTSPXVPYVAVASXAXHACQRRRLVD
GMSWMAIHGYNHPRLANAKAQT EOMSHVME DGI THQPAVDLCRRXVAMTPDALEC

```

```

VLADSGVAXEXVDENGASTAPLADRPPEYHXAFFLPXLO"
525..>965
/gene="biob"
525..>965
/gene="biob"
/note="similar to biob"
/codon_start=1
/product="biotin synthase"
/protein_id="AAG60564.1"
/db_xref="GI:12620106"
/translaction="MXHHARWMSOYTELFNKPELIMEFAQOVROHFPDRHOXST
LUSITGACPEDECKYCPQOSARKYTGLESRLMEYDVSARKAKNASTRCAGAA
KKPHIVTCPYLEOMWKGVEKMGLEALYDPRFAERRAXXSACLLA"
BASE COUNT 230 a 256 c 285 g 180 t 14 others
ORIGIN
Query Match 74.4%; Score 18.6; DB 1; Length 965;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 gcgcagcagtcgcgcgcgcatt 25
Db 603 GCCCAGACGTCATCGTCAGCACT 627

```

```

RESULT 31
LOCUS AF289025/c 1359 bp mRNA PLN 20-AUG-2000
DEFINITION Cucumis sativus plasma membrane H+-ATPase mRNA, partial cds.
ACCESSION AF289025
VERSION AF289025.1 GI:9858169
KEYWORDS
SOURCE Cucurbit.
          Cucumis sativus
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
          1 (bases 1 to 1359)
          Hub,S.M., Kim,S.M., Yang,Y.I. and Bai,D.G.
          Plasma membrane H+-ATPase partial cDNA of cucumber root
          unpublished
          2 (bases 1 to 1359)
          Hub,S.M., Kim,S.M., Yang,Y.I. and Bai,D.G.
          Direct Submission
          Submitted (23-JUL-2000) Department of Biological Sciences, College
          of Natural Sciences, Chonnam National University, 300 Yong Bong
          Dong, Pukgu, Kwangju 500 - 757, Korea
          Location/Qualifiers
FEATURES
            source
            1..1359
            /organism="Cucumis sativus"
            /db_xref="taxon:3659"
            /tissue_type="root"
            /note="Cucumis sativus L"
            <1..>1359
            /codon_start=1
            /product="plasma membrane H+-ATPase"
            /protein_id="AAG01028.1"
            /db_xref="GI:9858170"
            /translation="KRLGFMWNPISWMEAAATMAITLANGGKPPDMEDFGIICL
            VINSTSTIEENNAGNANAALMAGLAKTKYLRDGKSEDEEALIVPGDIVSKIDIL
            IPADARLLEGDPLKVDOSALTGESLPVTKNPGDEIFSGSTCKGGEIAVYIATGVTF
            FGKAHLVDSTHVGHFQKYLITAGNFCISIALGMLITIGMCPIDHRVYRGIDNL
            FVLLIGIPILAMPVLSVTMAIGSHKLSOGATIKRMJTALIEAGMVDVLCSDTGTGLT
            LNKLSVDNVLVEFAKGVQKHVILLARSASRTENOPADIAAIVGMADPKERARAGIR
            EVNFPFPPNPKRTALTYTIDSGNMWHSAGAPROITLNCNKEDYVKKAHAYIVDKA
            ERGLRSLAVGQEVPEKREKSPSPWQFVGLPLFDPDRDSEGTIKRALNLGVNVM
            ITG"
BASE COUNT 373 a 252 c 338 g 396 t
ORIGIN

```

	Query Match	Score	DB	Length
Best Local Similarity	74.4%	18.6	8	1359
Matches	21	Conservative	0	Mismatches 4
				Indels 0
				Gaps 0
Qy	1	gcgcagcaggtgcatcgccagcatt	25	
Db	193	gagcagcagcagcagcattgcccagcatt	169	

RESULT	32
STLDPHAL/c	
LOCUS	STLDPHAL
DEFINITION	S. tuberosum L. (Desiree) PHAI mRNA.
ACCESSION	X76536
VERSION	X76536.1 GI:435002
KEYWORDS	ATPase; phl gene; proton pump; transmembrane protein. potato.
SOURCE	Solanum tuberosum
ORGANISM	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE	1 (bases 1 to 3162)
AUTHORS	Harms,K.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1993) K. Harms, Institut fuer Genbiologische Fors. GmbH, Ihnestrasse 63, 14195 Berlin, FRG
REFERENCE	2. (bases 1 to 3162)
AUTHORS	Harms,K., Wohner,R.V., Schulz,B. and Frommer,W.B.
TITLE	Isolation and characterization of P-type H ⁺ -ATPase genes from potato
JOURNAL	Plant Mol. Biol. 26 (3), 979-988 (1994)
MEDLINE	95093037
FEATURES	
source	Location/Qualifiers I..3162

BASE COUNT	845 a	592 c	795 g	930 t
ORIGIN				

Query Match	74.4%;	Score 18.6;	DB 8;	Length 3162;
Best Local Similarity	84.0%;	Pred. No. 9e+02;		
Matches	21;	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0

```
QY      1 ggcagcaggtgcatgccagcatt 25
          | ||||| | ||||| ||||| ||
Db     408 GAGCAGCTGCTGCATTGCCAGCATT 38
```

RESULT	33				
LOCUS	AF029256/c				
DEFINITION	AF029256	3175 bp	mRNA	PLN	25-FEB-2000
ACCESSION	Kosteletzkya virginica KVAR1			plasma membrane proton ATPase (ATP1)	
VERSION	AF029256				
KEYWORDS	AF029256.2	GI:7105716			
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Cook,D.A.				
TITLE	A Sodium Chloride Induced Proton ATPase from a Salt Marsh Halophyte				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3175)				
AUTHORS	Cook,D.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-OCT-1997)				
COMMENT	University, 223 James P. Bralley Dr. SW, Atlanta, GA 30314, USA				
FEATURES	On Feb 25, 2003 this sequence version replaced gi:2605906.				
Source	Location/Qualifiers				
	1..3175				

BASE COUNT	878 a	618 c	795 g	884 t
ORIGIN				

```

Query Match      74.4%      Score 18.6; DB: 8; Length 3175;
Best Local Similarity 84.0%      Pred No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 ggcgagcaggtgcatcgccagcatt 25
          | | | | | | | | | | | | | | |
Db      407 GAGGAGCGAGCTGCATTCGCCAGCATT 383

```


RESULT 34
LOCUS TOBPM3A/C
DEFINITION Nicotiana plumbaginifolia plasma-membrane H+ ATPase (pma3) mRNA,
complete cds.
ACCESSION M80490
VERSION M80490.1
KEYWORDS GI:170293
SOURCE Nicotiana plumbaginifolia (library: lambda-gtl1) cDNA to mRNA.
ORGANISM Nicotiana plumbaginifolia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 3224)
AUTHORS Perez, C., Michellet, B., Ferrant, V., Bogaerts, P. and Boutry, M.
TITLE Differential expression within a three-gene subfamily encoding a
plasma membrane H⁺-ATPase in Nicotiana plumbaginifolia
JOURNAL J. Biol. Chem. 267, 1204-1211 (1992)
MEDLINE 92112762
FEATURES
source
location/Qualifiers
1..3224
/organism="Nicotiana plumbaginifolia"
/db_xref="taxon:4092"
/issue_lib="lambda-gtl1"
6..23
/note="upstream short ORF"
/codon_start=1
/protein_id="AAA34097.1"
/db_xref="GI:170294"
/translation="MFLI"
66..2936
/gene="pma3"
/gene="pma3"
66..2936
/gene="pma3"
/product="plasma membrane H⁺ ATPase"
/codon_start=1
/protein_id="AAA34098.1"
/db_xref="GI:170295"
/translation="MGEKPEVLDVAVLKEVDLENPIEEVEFENLCTREGSLRTATAOE
RLSIFGYNKLEEKESKESKFLGFMNPLSWVEAAIMATIALANGGKPPMDQFVG
ITLLINSTISFTEENNAAGAAALMARLAKVLRGDKWKEDEAVLVPDGIISI
KIGDIIPADARLLEGGDLKIDGALGSEIPVTKGPDGVSSTCKGGEIEAVVAT
GVHTEFGAAHLVDSTNOVGHFOKYLTAIGNFCISIAVGMIEITIVVPIORHRYRP
GIDNLLVLLIGGPIAMPVLSVTMAISHRLAGQALTKRMTAIEEMAGMDVLCSDK
TGTLTLNKLTVDKYLEVFAVGVADGVLMHARSRENDAIDAAIYGMADPKEA
RAGIREIHFLPENTDKRTALTYLDGEGMHVSGAPEQLIHLAHNSDIERVHAV
IDKFAERGLRLAVAYOEVEGRKESAGPMOFLALPLDPPRHDSAEITERRALNG
VAVKMTGQDLATICKETGRGLGCMGTNNMPSALLGOTDESIAPVDELLERADGFA
GVPEHKEIVTRLOARKHITCGMTGCVDAALAKKADIGAVDDATDAARSADIVL
TEPGLSVITSAVLTSAITFQRMKNNTTAVSITIRIVLGFMLALIMQDFPPENYLI
IALNDGILMTISDKRVKPSPLPDSMKLAIEITTVGVIGYLAAMTVIFFWAAYTNF
PFRVGVSTLEKATDDEFRLASAIYLOVSTISQALIEVTRSRSPMERDGLLVFA
FFPAOLVATLIIVANMSFAALEGIGWAGVIMLVNIIVYIPLDXXFLIRALYLSG
AMDVLVORTIAFTPKRDKGKELRELQWNAHQRLLHGLQVDPKIFSETTNFNLNOLA
EFAARRAEIARLRELHTLKGHVSVYKLKGLDITETIQQVTV"
3224
polya_site
/gene="pma3"
BASE COUNT 837 a 607 c 821 g 957 t 2 others
ORIGIN

Query Match 74.4% Score 18.6; DB 8; Length 3224;
Best Local Similarity 84.0%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcagcaggtgcatcgccagcatt 25
||||| ||||| ||||| ||||| |||||
Db 447 GGGCAGCTGCTGCATTCGCCAGCAT 423

RESULT 35
LOCUS TOMLHA1/C

LOCUS TOMLHA1 3229 bp mRNA PLN 27-APR-1993
DEFINITION Tomato (L. esculentum) H⁺-ATPase (LHA1) mRNA, complete cds.
ACCESSION M60166
VERSION M60166.1
KEYWORDS GI:170463
SOURCE H⁺-ATPase; plasma membrane protein; transmembrane protein.
ORGANISM L. esculentum, cDNA to mRNA.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 3229)
AUTHORS Ewing, N.N., Wimmers, L.E., Meyer, D.J., Chetelat, R.T. and
Bennett, A.B.
TITLE Molecular cloning of tomato plasma membrane H⁺-ATPase
JOURNAL Plant Physiol. 94, 1874-1881 (1990)
MEDLINE 19071000
FEATURES
source
location/Qualifiers
1..3229
/organism="Lycopersicon esculentum"
/db_xref="taxon:4081"
121..2991
/gene="LHA1"
121..2991
/gene="LHA1"
/EC_number="3.6.1.35"
/product="H⁺-ATPase"
/protein_id="AAA34173.1"
/db_xref="GI:170464"
/translation="MAEKPEVLDVAVLKEVDLENPIEEVEFENLCTREGSLRTATAOE
RLSIFGYNKLEEKESKESKFLGFMNPLSWVEAAIMATIALANGGKPPMDQFVG
ITLLINSTISFTEENNAAGAAALMARLAKVLRGDKWKEDEAVLVPDGIISI
KIGDIIPADARLLEGGDLKIDGALGSEIPVTKGPDGVSSTCKGGEIEAVVAT
GVHTEFGAAHLVDSTNOVGHFOKYLTAIGNFCISIAVGMIEITIVVPIORHRYRP
GIDNLLVLLIGGPIAMPVLSVTMAISHRLAGQALTKRMTAIEEMAGMDVLCSDK
TGTLTLNKLTVDKYLEVFAVGVADGVLMHARSRENDAIDAAIYGMADPKEA
RAGIREIHFLPENTDKRTALTYLDGEGMHVSGAPEQLIHLAHNSDIERVHAV
IDKFAERGLRLAVAYOEVEGRKESAGPMOFLALPLDPPRHDSAEITERRALNG
VAVKMTGQDLATICKETGRGLGCMGTNNMPSALLGOTDESIAPVDELLERADGFA
GVPEHKEIVTRLOARKHITCGMTGCVDAALAKKADIGAVDDATDAARSADIVL
TEPGLSVITSAVLTSAITFQRMKNNTTAVSITIRIVLGFMLALIMQDFPPENYLI
IALNDGILMTISDKRVKPSPLPDSMKLAIEITTVGVIGYLAAMTVIFFWAAYTNF
PFRVGVSTLEKATDDEFRLASAIYLOVSTISQALIEVTRSRSPMERDGLLVFA
FFPAOLVATLIIVANMSFAALEGIGWAGVIMLVNIIVYIPLDXXFLIRALYLSG
AMDVLVORTIAFTPKRDKGKELRELQWNAHQRLLHGLQVDPKIFSETTNFNLNOLA
EFAARRAEIARLRELHTLKGHVSVYKLKGLDITETIQQVTV"

BASE COUNT 833 a 616 c 813 g 947 t
ORIGIN Chromosome 3.

Query Match 74.4% Score 18.6; DB 8; Length 3229;
Best Local Similarity 84.0%; Pred. No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggcagcaggtgcatcgccagcatt 25
||||| ||||| ||||| ||||| |||||
Db 502 GGGCAGCTGCTGCATTCGCCAGCAT 478

RESULT 36
LOCUS TOBPM3B/C
DEFINITION Nicotiana plumbaginifolia plasma-membrane H⁺ ATPase (pma3) gene,
exons 1-9.
ACCESSION M80491
VERSION M80491.1
KEYWORDS GI:170296
SOURCE Nicotiana plumbaginifolia (library: EMBL4 genomic) DNA.
ORGANISM Nicotiana plumbaginifolia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 3233)

AUTHORS Perez, C., Michelet, B., Ferrant, V., Bogaerts, P. and Boutry, M.
 TITLE Differential expression within a three-gene subfamily encoding a
 JOURNAL plasma membrane H⁺-ATPase in *Nicotiana glauca* encoding a
 MEDLINE J. Biol. Chem. 267, 1204-1211 (1992)
 FEATURES 92112762
 SOURCE Location/Qualifiers

1..3233
 /organism="Nicotiana glauca" [ta] "Nicotiana glauca"
 /db_xref="taxon:4092"
 /tissue="leaf"
 /tissue_id="EMBL4 genomic"
 join(578..895,988..1107,1636..1734,1826..1963,2102..2236,
 2325..2504,2605..2811,2906..3025,3122..3233)
 /partial
 /gene="pma3"
 /product="Plasma membrane H⁺ ATPase"
 578..895
 /number=1
 /gene="pma3"
 /number=1
 782..799
 /note="upstream short ORF"
 /codon_start=1
 /protein_id="AAA34100.1"
 /db_xref="GI:170298"
 /translation="MVFLL"
 join(842..895,988..1107,1636..1734,1826..1963,2102..2236,
 2325..2504,2605..2811,2906..3025,3122..3233)
 /gene="pma3"
 join(842..895,988..1107,1636..1734,1826..1963,2102..2236,
 2325..2504,2605..2811,2906..3025,3122..3233)
 /partial
 /gene="pma3"
 /codon_start=1
 /product="Plasma membrane H⁺ ATPase"
 /protein_id="AAA34099.1"
 /db_xref="GI:553114"
 /translation="MGEKPEVLDVAVLKETVDLENIPLEVEFENLRCTREGITATAOE
 RLSTFGYNKLEEKESKFLKFLGFMNPLSMWEAAINAIALANGGKPPDQDFVG
 IITLLIINSTISFENNAGNAALAPARAPKAVLRDGMKEEDAVIVPDIIISI
 KLGDIIIPADARLEGDPLKIDOSALTGESLPTKPGDGYSGSTKOGIEEVAVAT
 GVHTFEGKAHLVSTNQVGHFQKVLTAIGNFCISAVGMILIIIVMYICIQRRKP
 GIDNLVLLIGGIPAMPVYLSVTMAIGSRILAOQCAITRMTVAIEEMAGMDVCSCK
 TGITLTKLTVDKYLIEVFARGVADTVVLMARASRTENQDAIDAATVGM.LAD"
 896..987
 /gene="pma3"
 /number=1
 988..1107
 /gene="pma3"
 /number=2
 1108..1635
 /gene="pma3"
 /number=2
 1636..1734
 /gene="pma3"
 /number=3
 1735..1825
 /gene="pma3"
 /number=3
 1826..1963
 /gene="pma3"
 /number=4
 1964..2101
 /gene="pma3"
 /number=4
 2102..2236
 /gene="pma3"
 /number=5
 2237..2324
 /gene="pma3"
 /number=5
 2325..2504
 /gene="pma3"
 /number=6
 2505..2604

BASE COUNT 901 a 542 c 636 g 1154 t
 ORIGIN
 Query Match 74.4% Score 18.6; DB 8; Length 3233;
 Best Local Similarity 84.0% Pred. No. 8.9e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 ggcgacgagtgatcgccacgacatt 25
 ||||| ||||| ||||| ||||| |||||
 Db 1934 GGCACGCTGCTGCATTCGCACGATT 1910
 RESULT 37
 AF275745/c 3244 bp mRNA PLN 10-AUG-2000
 LOCUS
 DEFINITION Lycopersicon esculentum plasma membrane H⁺-ATPase (LHA2) mRNA,
 complete cds.
 ACCESSION AF275745
 VERSION AF275745.1 GI:9789538
 KEYWORDS
 SOURCE
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Solanales; Solanaceae; Solanum;
 Lycopersicon
 1. (bases 1 to 3244)
 Ewing, N.N. and Bennett, A.B.
 Assessment of the number and expression of P-type H⁺-ATPase genes
 in tomato
 JOURNAL Plant Physiol. 106 (2), 547-557 (1994)
 MEDLINE 95083753
 REFERENCE 2 (bases 1 to 3244)
 Ro, S., Kirschke, C.P. and Ewing, N.N.
 Tomato plasma membrane H⁺-ATPase LHA2 plays a role in
 auxin-regulated processes
 JOURNAL unpublished
 REFERENCE 3 (bases 1 to 3244)
 Ro, S. and Ewing, N.N.
 Direct Submission
 Submitted (06-JUN-2000) Biological Sciences, California State
 University, Sacramento, 6000 J Street, Sacramento, CA 95819-6077,
 USA
 FEATURES
 SOURCE Location/Qualifiers
 1..3244
 /organism="Lycopersicon esculentum"
 /cultivar="Alisa Craig"
 /db_xref="taxon:4081"
 /chromosome="6"
 /map="between CSC83 and TGI62"
 1..3244
 /gene="LHA2"
 94..2964
 /gene="LHA2"
 /note="P-type ion pump"

Query Match	74.4%	Score 18.6;	DB 8;	Length 3244;
Best local similarity	84.0%;	Pred. No. 8.9e+02;		
Matches 21;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
Oy	1	ggcgagcaggtgcatcgacgacatt	25	
Db	475	GAGCAGCTGCTGCATGCGACGACTT	451	
RESULT 38				
LOCUS	MCU84891/c			
DEFINITION	Mesembryanthemum crystallinum H+-transporting ATPase (pMA) mRNA,			
ACCESSION	MCU84891			
VERSION	084891			
KEYWORDS	complete cds.			
SOURCE	084891.1			
ORGANISM	GI:1814406			
	common ice plant.			
	Mesembryanthemum crystallinum			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.			
REFERENCE	1 (bases 1 to 3426)			
AUTHORS	Michalowski,C.B., Quigley-Landreau,F., and Bohnerl,H.J.			
JOURNAL	Mesembryanthemum crystallinum cDNA for H+-transporting ATPase			
REFERENCE	unpublished			
AUTHORS	2 (bases 1 to 3426)			
TITLE	Michalowski,C.B., Quigley-Landreau,F., and Bohnerl,H.J.			
JOURNAL	Direct Submission			
	Submitted (13-JUN-1997) Biochemistry, University of Arizona, Bio			
	Sciences West 513, Tucson, AZ 85721, USA			
FEATURES	Location/Qualifiers			
source	1..3426			
	/organism="Mesembryanthemum crystallinum"			
	/db_xref="taxon:3544"			
	227..3088			
gene	/gene="pMA"			
	227..3088			
CDS	/gene="pMA"			
	/EC-number="3.6.1.35"			
	/function="plasma membrane proton pump"			
	/codon_start=1			
	/product="H+-transporting ATPase"			
	/protein_id="AAB41898.1"			
	/db_xref="GI:1814407"			
	/translation="MVGNSGSLQEIKEDEVDLEKIPTEFVPSLKSREGLSAEEG			
	ANRIQFGPNKLEEKDKSLKPIGFMNPLSWMEAAALMAIVLANDANCPMDPMD			
	VGITLLVINSTISFTEENMGNAALMANLAPRTKLRGCRNGECSALIVGDI			

BASE COUNT	951	a	643	c	864	g	968	t
ORIGIN								
Query Match	74.4%: Score 18.6; DB 8; Length 3426;							
Best Local Similarity	84.0%: Pred. No. 8.9e+02;							
Matches	21; Conservative	0; Mismatches	4; Indels	0; Gaps	0; Mismatches	4; Indels	0; Gaps	0; Mismatches
QY	1	gcccagcagatgcatgcacgacatc 25						
Db	614	GGCAGCGCTGCATTACCGACATP 520						
RESULT 39								
AF179442/c								
LOCUS								
DEFINITION	AF179442 10302 bp DNA PLN 08-JUN-2000							
	(Lycopersicon esculentum plasma membrane H+-ATPase isoform LHA2 (LHA2) gene, complete cds; and 18S ribosomal RNA gene, partial sequence.							
ACCESSION	AF179442							
VERSION	AF179442.1							
KEYWORDS	GI:5901756							
SOURCE	tomato.							
ORGANISM	Lycopersicon esculentum							
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;							
REFERENCE	Lycopersicon.							
AUTHORS	1 (bases 1 to 10302)							
	Ewing,N.N., Wimmers,L.E., Meyer,D.J., Chetelat,R.T. and Bennett,A.B.							
TITLE	Molecular cloning of tomato plasma membrane H+-ATPase							
JOURNAL	Plant Physiol. 94, 1874-1881 (1990)							
REFERENCE	2 (bases 1 to 10302)							
AUTHORS	Ewing,N.N. and Bennett,A.B.							
TITLE	Assessment of the number and expression of P-type H(+)-ATPase genes							
JOURNAL	in tomato							
MEDLINE	Plant Physiol. 106 (2), 547-557 (1994)							
PUBMED	95083753							
REFERENCE	7991683							
AUTHORS	3 (bases 1 to 10302)							
TITLE	Ro.S., Idate,R.R. and Ewing,N.N.							
JOURNAL	The promoter of plasma membrane H+-ATPase LHA2 directs reporter							
REFERENCE	gene expression in Arabidopsis in a pattern that suggests a role							
AUTHORS	for this isoform in auxin-regulated processes							
TITLE	Unpublished							
JOURNAL	4 (bases 1 to 10302)							
REFERENCE	Ro.S., Idate,R.R. and Ewing,N.N.							
AUTHORS	Direct Submission							
TITLE	Submitted (22-AUG-1999) Biological Sciences, California State							
JOURNAL	University, Sacramento, 6000 J Street, Sacramento, CA 95819-6077,							
USA								
FEATURES	Location/Qualifiers							
Source	1..10302							
	/organism="Lycopersicon esculentum"							
	/cultivar="VFN Cherry"							
	/db_xref="taxon:4081"							
	/chromosome="6"							
	/map="between CSC83 and TG162"							

promoter 1.2146
/gene="LHA2"
gene 1.9460
/gene="LHA2"
mRNA join(2147..2218,2347..2421,2517..2636,3006..3104,3191..3328,3475..3609,3700..3879,3976..4182,4302..4421,4505..4624,4705..4827,4918..5022,5229..5373,5592..6156,6269..6507,6694..6726,6817..6898,7480..7640,8113..8298,8379..8561,8646..8824,9084..9460)
/gene="LHA2"
/product="plasma membrane H+-ATPase isoform LHA2"
/evidence=experimental
join(2368..2421,2517..2636,3006..3104,3191..3328,3475..3609,3700..3879,3976..4182,4302..4421,4505..4624,4705..4827,4918..5022,5229..5373,5592..6156,6269,6694..6726,6817..6898,7480..7640,8113..8298,8379..8561,8646..8824,9084..9180)
/gene="LHA2"
/codon_start=1
/product="plasma membrane H+-ATPase isoform LHA2"
/protein_id="AAD55399.1"
/db_xref="GI:5901757"
/translation="MGKPEVLDAVLEKFTVDLENIPLEEVFNLRCTKEGLTGTAAQRIATIGYKMLEEKESKFLKEFGMMNPLSWMEAAALMALANLGGGKPPDMDFVGIITLLVINSTISLEENNAGNAALMALAPKAVLRDKNEDAAVLPGDIISI KLGGIIPADARLEEDPLKIDQSLGTESLPTVKGPGGVSGSTCKOGELEAVIATGVHPFGAAHLVDSTNOVGHOKVLTALGNPCISLAVGMITFIIWYPIORHKYRPGDNLVLLIGSIPAMPTVLSYMAISGSHLAOOGATTKRMTAIEEAGMDVLCISKTCETLNLKLTVDKNI LVEYFAKGVADIVYLMARASRTENDATIDTALVGM,SDPKERAGIREIHELPNPPTDKRTALTYLDEGKMRVSKAPEQILNLANKSDIERVHAVIDKEERGLRSIGAVYQVEPEKREKSSGPMOITGLPLPDPHSDSETIRALNLGVNWKMTGDOLAIGKETGRRLGMGTNMPSSALGOTKDESIASLPIDELIEKADGFAVPEPHKEIYKRLQARKHICGMGTGYNADAPALKKADIGIANDATDAASASDIULTPEGSIVTISAVLSRAIFORMKNTTYVANSITIRIVYGMILLAIKPFDPPEPMVLI TALINDGTIMTISDRKPSPLPSMKLAETFTTGVYUGGLAMMTVIFPMAIETDLPFRVGVSTLORTATDPRKASATYLVOSTISQALPFTRSRMSFEVERGLLEVAFLIADLVATLAVASWSFAIEIGIWMGAGVIMLYNLYEFPLDIITKFLIRVALSKRAMDVLEQRIAFTRKRKDGKEDRELQWMAHQRTLHGLOVPDTKLFSEATNNEINQALBEAKRKAIARQRELHLKGHVESVYKLKGDIEIIOOSYTV"
tRNA /product="18S ribosomal RNA"
BASE COUNT 3040 a 1596 c 1989 g 3676 t 1 others
ORIGIN

Query Match 74.4% Score 18.6; DB 8; Length 10302;
Best Local Similarity 84.0% Pred. No. 7.9e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcgcagcagtcgacgcgcagcatt 25
1 ||||| ||||| ||||| ||||| |||||
Db 3299 GAGCAGCTGCTGATGCCGCATTT 3275

RESULT 40 AE008283 14379 bp DNA BCT 14-AUG-2001
LOCUS AE008283/C Agrobacterium tumefaciens strain C58 linear chromosome, section 87
DEFINITION of 187 of the complete sequence.
ACCESSION AE008283 AE007870
VERSION AE008283.1 GI:15159272
KEYWORDS
SOURCE
ORGANISM Agrobacterium tumefaciens.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 14379)
Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), The Causative Agent of Crown Gall
Disease in Plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14379)

AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
Street, Cambridge, MA 02139, USA
FEATURES
source
gene
144..1136
/gene="AGR_L.1682"
/db_xref="taxon:358"
CDS
144..1136
/gene="AGR_L.1682"
/note="MTO-INDUSTOL 2-DEHYDROGENASE"
/codon_start=1
/transl_table=11
/product="AGR_L.1682p"
/protein_id="AK89414.1"
/db_xref="GI:15159273"
/translation="WYTKLALGAGRIGVHAKAITTDSRAKLAVAVDAMAGABAIA
REAQCVSTIDEVLADKTIDAVITCTPTNTHADLIERFARAKAIFCEKPYDLVARA
EACAKVEETGAKVMGLGNRPDPHFOAVHKAIEDRLGKVMVITTSHPDPPAEY
IKVSGIIFRDMTTHDFDMARFLGEEVESVPATGSYLVNPEIGALGDPPSASIIITTK
SGROALINSRRTYGDORIEVHSGSAAENORPVSIFIANNDGTPTPLHPFM
TRYTAVANRIAFIDISINNPSPTSIDGLIALKLADALKSATSKTVITIG"
1495..2379
/gene="AGR_L.1685"
1495..2379
/gene="AGR_L.1685"
/note="hypothetical protein PA0239 {imported} -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L.1685p"
/protein_id="AK89415.1"
/db_xref="GI:15159274"
/translation="MPSAKGFLVALVAFITFAODGTSKHLGSAPPVYAMLRYWA
FAFVFLMARSGLICGHRKCAIFVGTIGVYILKPPDDGDKNLIITTFVAM
ASAPLIVAAISMLGICGHRKCAIFVGTIGVYILKPPDDGDKNLIITTFVAM
LAIYSLTRGKSGSDSASMTSFTYGVAGAAULTLPYFWPLAPDGMKMLATITFG
MSGHVCLIAFELADAAVQPSYVQYLVLSITGVITGEVNTSMKVGAIIVIAAGL
FTIHRHVAVARRNRREAN"
complement(2376..2792)
/gene="AGR_L.1687"
complement(2376..2792)
/gene="AGR_L.1687"
/note="(U39262) ornithine cyclodeaminase"
/codon_start=1
/transl_table=11
/product="AGR_L.1687p"
/protein_id="AK89416.1"
/db_xref="GI:15159275"
/translation="MYGEGITHNAVGGDCPKTELMRLILRSDFIYVPPOTRBE
IOOLPADVPYELKQVYISGKQGRAGDQITLFDVSGFATEDFSALRVRSKLATGL
YTELIDLADPDPRDLFGMLLRCEALVQTDGROA"
2940..4478
/gene="AGR_L.1690"
2940..4478
/gene="AGR_L.1690"
/note="possible transposase of IS1353"
/codon_start=1
/transl_table=11
/product="AGR_L.1690p"
/protein_id="AK89417.1"
/db_xref="GI:15159276"
/translation="MYSVADRLRAVELYIRLGRKLNITIROLGPTKNAALGWREY
GHILDRTPVAPAKPYSEAKORALIEHRTHDRKISGTRALGTPGRTLAVAREAP
PEARLSWGRSWHPQISEVROAGVIGLCSDESQVADLGSVPGLYSWKQDLG
HEASSSKRRKANRPVPERBELERKLEALQIARSSYFYBARMCADKAAVRLSLAET
LILSNREKTOILDALKEVYRLPELLAQIARSSYFYBARMCADKAAVRLSLAET
FEANRRGYRLOSLARKSVYISEKVVORLMOEHLAVARPRRRGSGYLGEISPA
PENLIRPDHADAPVWKLTIDITEROIPAGKVYISPTDCPDGAVISGTSCTGODACL
VNTMLDAITGVANGERPILHSRGAHYRMPGWLTRISEARLVRSRSGCSODNAA

CEGFGRLKTELEYPDWMKAITIEQVAVEDVAYIRWYNEKRIKISLGSLSPEYKOSL
GLKL"

gene
/gene="AGR_L_1691"
complement(4493..5065)
/note="(U39262) ornithine cyclodeaminase"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1691p"
/protein_id="AAK89418.1"
/db_xref="GI:15159277"
/translation="MTVLPNLTNPVPEVSDHMKLVKVGIDTFLRELAEVEEDRR
WQSDKTPRIASHKEGVIFLIMPTSDGTLGFKYNGHPKPKMKEGQTVYAFGLSDV
GNGYPLLTETITLTATATSAVAKYIAPARSMALIGNGAQSEFOARAFKAL
GIDTFLRFDIDPSSEKCARNLAQGGCCG"

gene
complement(5079..6062)
/gene="AGR_L_1693"
/note="(U39262) arginase"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1693p"
/protein_id="AAK89419.1"
/db_xref="GI:15159278"
/translation="MFLPCLHGLVORSVCKARTGMDIRLVGAPLQIGAGQCGEMSP
SAYVAGLAHLEBELGHVVDTCNMVMPARECPNPVHILAEIVAMTEALTEAV
RESADVAFLGSDHAIASGTAVAGMARVATGPFVLMDATDVHITETTSGLM
HGTAVPFGSDGSGVSPPLSHAVAEENIGMIGISVDPAERALEKSGITVHDMSS
IDEGVAVITREPLARVOAAGLLHVSIDVDLEPSTIAPAGTTPGCGATFERHLYM
EMHDSGLVCSLDLVELNPLFDEKRTATLMDVLTSLMKRWDRPTKAC"

gene
/gene="AGR_L_1694"
/note="(U39262) arginase"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1694p"
/protein_id="AAK89420.1"
/db_xref="GI:15159279"
/translation="MDTIDFRLITLRNCRSTSDLAETETSRATVRRATKEMND
GSIIGYVILRADAEANRIGTMTETGHTDRTYIRLGGFPELSEIHITNGMDLI
VELNAATLSDPDVALRIRLIVPGITGSETSLSTPRSTRAL"

gene
complement(6823..7665)
/gene="AGR_L_1696"
/note="(U39262) arginase"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1696p"
/protein_id="AAK89421.1"
/db_xref="GI:15159280"
/translation="MTDESVRFLTSLVELGSLTAARAGTVQPVVSQIKLMLEQRL
GRRLLDRSPRLRLTDEGQVLPKARELLAKHEDALNFSDEPPLHFAALASDHAI6SR
LSAAILRAMKASIPANAVIDVTGFSRQVKESSAGIADAIIRRNQSGSDEVELPTDR
LGMVAPDWRPRANSAPVLSLPGCGVDIAISELQAGVYWRDAFTAGSCSALLEG
VHAGAGTALGDISARGLPDRGAEELGPELPSPDVLLSRATPASHA1R1ALVAAS
GAGD"

gene
/gene="AGR_L_1698"
/note="(U39262) arginase"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1698p"
/protein_id="AAK89422.1"
/db_xref="GI:15159281"

translation="MMSAOGFROGGDDTMPMIKARYTRPONTADPKLEVALLSRL
TAEITHKDPATVAVLIVESADPADMTIASSVAEOKVATFYEVSTHTDNTNKDTEEM
VARYVEENRRLGPVHEASVYLVAHVSDADATGGOOTONARKHV"

gene
/gene="AGR_L_1699"
complement(8202..9095)
/note="(U39262) ornithine cyclodeaminase"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1699p"
/protein_id="AAK89423.1"
/db_xref="GI:15159282"
/translation="MTVLPNLTNPVPEVSDHMKLVKVGIDTFLRELAEVEEDRR
WQSDKTPRIASHKEGVIFLIMPTSDGTLGFKYNGHPKPKMKEGQTVYAFGLSDV
GNGYPLLTETITLTATATSAVAKYIAPARSMALIGNGAQSEFOARAFKAL
GIDTFLRFDIDPSSEKCARNLAQGGCCG"

gene
complement(9132..10217)
/gene="AGR_L_1702"
/note="(AF311738) B102"

CDS
/codon_start=1
/transl_table=11
/product="AGR_L_1702p"
/protein_id="AAK89424.1"
/db_xref="GI:15159283"
/translation="MRLPGLSLPMSHRONTASANGDVOTIAKVLTMOTRSSRAGF
GHVAPRCVDAEIEASLGEAGMIRRTGISRYVAEAGDTLSGLAERAGMALEDA
KINADDIATLITATSTPDHILPSPARLHRLGTLTSGAIDLAGACGSLVLTADG
FVRTYGRVAVYVANILSRINPAPASVITPDAGAVITPCPEKRGVLSADVA
DSSGVDLIQIAGGSSOPFSAGNTADALMTDRDGEVSRVATMTNOSQVLEHAE
LTAADISREYHQANARMSDQAVGNGICREKTVRTISGNSAATPLSLITNAE
RPLACGETLLTAAAGAMTGCAVAVYRV"

Query Match 74.4% Score 18.6; DB 1; Length 14379;
Best local similarity 84.0%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcatcgcaagcatt 25
|||||
DB 14118 GCGCAGCAGGTGCAACCGCTGCATT 14094

RESULT 41
AC017995/33204 bp DNA HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
DEFINITION pieces
AC017995
VERSION AC017995.1 GI:6553195
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 33204)
AUTHORS Adams, M. and Venter, J.C.
TITLE Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDL-10212937 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source location/Qualifiers
1..33204

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT      8981 a      7720 c      7703 g      8800 t
ORIGIN
Query Match      74.4%; Score 18.6; DB 2; Length 33204;
Best Local Similarity 84.0%; Pred. No. 7,1e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ggcagcaggtgcagtcagcagcat 25
      ||| ||| ||| ||| ||| |||
Db 3374 GCCTACGACCTGCACCGCCACCAT 3350

RESULT 42
LOCUS  ATE27H5      88548 bp      DNA      PLN      14-Apr-2000
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone F27H5.
ACCESSION AL163852
VERSION AL163852.1 GI:7576197
KEYWORDS
SOURCE
ORGANISM
        thale cress.
        Arabidopsis thaliana
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
        Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
        1 (bases 1 to 88548)
        Rieger,M., Mueller-Auer,S., Zipp,M., Schaefer,M., Mewes,H.W.,
        Rued,S., Lemcke,K., Mayer,K.F.X., Quettler,F. and Salanoubat,M.
        unpublished
        2 (bases 1 to 88548)
        EU Arabidopsis sequencing project.
        Direct Submission
        Submitted (14-APR-2000) MIPS, at the Max-Planck-Institut fuer
        Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
        lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
        Coordinator: Marcel Salanoubat and Francis Quettler, Grouperment
        d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
        Gaston Creneau, BP191, 91006 Evry Cedex, France;
        http://www.genoscope.cns.fr
        Information on performance of analysis and a more detailed
        annotation of this entry and other sequences of chromosomes 3, 4
        and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
        source
            1..88548
                /organism="Arabidopsis thaliana"
                /variety="Columbia"
                /db_xref="taxon:3702"
                /chromosome="3"
            1..28386
                /note="This BAC overlaps with BAC T209. Please refer to
                EMBL accession AL138658 for analysis and annotation."
                /complement(29790..30794)
                /gene="F27H5_10"
                /gene="F27H5_10"
                29790..30794
                    /gene="F27H5_10"
                    /complement(29790..30794)
                    /note="A probable sequencing error has led to the change
                    in ORF which changes the first 5 amino acids (compared to
                    Ap132014). Genoscope has been asked to check."
                    /codon_start=1
                    /product="RING-H2 zinc finger protein ATL4"
                    /protein_id="CAB87861.1"
                    /db_xref="GI:7576198"
                    /translation="MESLNPISHGGGNDSSSSIDSLKPSVLIIVIIILMLIISVS
                    ICLRLNRCSHRSVLPISSSSSSVATVTSDRRSGCHPVSFETERSVLSLPIFK
                    SSYTRSSSMNSGDCAVCSKPEPEQRLILPLCCAHADCTIDMLVSNQCPICRS
                    PLFASPSDLKSLAVYGSNNGGGNSFRLETSISRSRTPPIPESEVQHRYSISFD
                    YIVDVDSLETSNFRKQEDFTTTATATATVATNPISFEASLAADIGNDSRSLK
                    DVYRLSRGISSKAMSFRSSGRFTGSSSRSEELIVMDLEAHAGEEISELFRMLSGV
                    "
        misc_feature
            /note="This BAC overlaps with BAC T209. Please refer to
            EMBL accession AL138658 for analysis and annotation."
            /complement(29790..30794)
            /gene="F27H5_10"
            /gene="F27H5_10"
            29790..30794
                /gene="F27H5_10"
                /complement(29790..30794)
                /note="A probable sequencing error has led to the change
                in ORF which changes the first 5 amino acids (compared to
                Ap132014). Genoscope has been asked to check."
                /codon_start=1
                /product="RING-H2 zinc finger protein ATL4"
                /protein_id="CAB87861.1"
                /db_xref="GI:7576198"
                /translation="MESLNPISHGGGNDSSSSIDSLKPSVLIIVIIILMLIISVS
                ICLRLNRCSHRSVLPISSSSSSVATVTSDRRSGCHPVSFETERSVLSLPIFK
                SSYTRSSSMNSGDCAVCSKPEPEQRLILPLCCAHADCTIDMLVSNQCPICRS
                PLFASPSDLKSLAVYGSNNGGGNSFRLETSISRSRTPPIPESEVQHRYSISFD
                YIVDVDSLETSNFRKQEDFTTTATATATVATNPISFEASLAADIGNDSRSLK
                DVYRLSRGISSKAMSFRSSGRFTGSSSRSEELIVMDLEAHAGEEISELFRMLSGV
                "
        gene
            /complement(29790..30794)
            /gene="F27H5_10"
            /gene="F27H5_10"
            36842..37038
                /gene="F27H5_20"
                /number=1
                36842..37893
                    /gene="F27H5_20"
                    /number=1
                    /join(36842..37038,37144..37256,37378..37512,37638..37699,
                    37777..37893)
                    /gene="F27H5_20"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB87860.1"
                    /db_xref="GI:7576199"
                    /translation="MSYNOSRRDRETFQYRTGRSTGNQOQOQOHRSSSAAGYKGA
                    AGSAPAPSTYFDNSLSNSKSPKGNAGGGGPRVNLPPVNHNNHNNHNSRS
                    QGEPYGGPTNPTESFNRTGPIPAAPNSQSTVMSSKINERTPNTAKASGASQAFPLQ
                    FGSIGPDLNVPARTTSAPPNMDOKRAQVEHLLVICRYCPCCLHK"
                    37039..37143
                        /gene="F27H5_20"
                        /number=1
                        37144..37256
                            /gene="F27H5_20"
                            /number=2
                            37257..37377
                                /gene="F27H5_20"
                                /number=2
                                37378..37512
                                    /gene="F27H5_20"
                                    /number=3
                                    37513..37637
                                        /gene="F27H5_20"
                                        /number=3
                                        37638..37699
                                            /gene="F27H5_20"
                                            /number=4
                                            37700..37776
                                                /gene="F27H5_20"
                                                /number=4
                                                37777..37893
                                                    /gene="F27H5_20"
                                                    /number=5
                                                    38085..39449
                                                        /gene="F27H5_30"
                                                        /number=1
                                                        /join(38085..39449,39541..42284,42684..42878,42965..43227,
                                                        44009..44026,44321..44409,44508..44590,44687..44750)
                                                        /gene="F27H5_30"
                                                        /note="Similarity to eukaryotic protein synthesis
                                                        initiation factor, Homo sapiens, EMBL:AF104913"
                                                        /codon_start=1
                                                        /product="protein synthesis initiation factor-like"
                                                        /protein_id="CAB87861.1"
                                                        /db_xref="GI:7576200"
                                                        /translation="MTRANVLYOMQSSLTPTASVAPSVKPKSSKNGKADNLMKREG
                                                        HNPSEKADIVPHIAPPSTQKSPITNIRMPVSQTPYQHTQVPHVPHFGPPMNMOT
                                                        PVTAITSQMPMPMALSMGNTPOLIPPOVEYGHPRPHNHGMMHQAQGHGFAIPMGQ
                                                        IHPOLGHVGOVLSPOYPQOOGKYGARKTTPVLTIPDHHEELDLDRDDPYSEGS
                                                        TALPKHNSMPRPSQPVSAFAPRVNLQPSVNSMTMLYPPVSVPLNNGRSDPYSPY
                                                        HYVIDSGORVOLIINPAHATPOLIRPAAPAHLLSDSTSVKARNQWNSALPYA
                                                        KSVKPAVGSFRKSGSPKDRSGEYNTLSOKNNVACSLSSQOPKPSFYGCVPRNSAP
                                                        PAKSPVEPLAKSSVETVPKSVSTTAVTTTELIRRAVSESTISEEDQCKVEP
                                                        HNLTEKQHTPDSLVSDPETAVAAKENSLPATNGFRQLIKVSTTSAPTSQSDVD
                                                        TSDKSTEGSSHASSEISGSPQEDLCKDNKASDLDERSVISAKHETLSGVLEK
                                                        AONVEDATQDQVSEKILAVTDTSSDLPSTHYLSVTLGSHETKSAVENTRNL
                                                        TSTGKKRIKEIILQKADAGTSDLYAAKYGPEKKSSNVVHDVSNQNLPAIPAY
                                                        EAVDTPEVKNKEPDMEDADADVSPKLETDNSVYNAKRSQSDVSNQITTEKKYRD
                                                        FLKPADLCIALPQGPVSPDIANALIVAAWAGSHHEDSYPTFGKVMRQASGARLD
                                                        RRENVAAGDDBRWIKNOSLPAGTGVGAVFRPGGQNSGVILRNRMGCGPIISRMQVPG
                                                        PMGQGRNTDPLERMQGSNFPQKGLFSPHTPMQVHNAKERYQVGTIADBOAKOR

```

OLKSLINKLTPONFEKLEQVKSVDNIDAVLTSGVLSQIFDKALMEPTCEWADFC
 HSGALPDNENGEKIFKRLILNKCCEPERKEEESRAVEEQVOTEEREE
 KRIQVRRMLGNILIGELYKKRLTEKIMHACJOKLGNODPHEINIEALCKLEST
 IGVNIDHNKAKFOMDGYFEKKMKLSGCOELSSRYRPLJVAIDIRKMKOERKLVST
 KRIEIVHRDAOERQOTANLRSRGSGRGHMFSSRGGGGLSPPAOMOSY
 HGPPOGRFSSNODIRFDPRSPYEPMPHORSYCEPPTLPGGGLGCOGMSLRPVP
 ASNTYQSDATQAGGDSRRPAGLNGFGRHPASPVTHGRSSFOERGTAVVHEPAST
 SRASDLSPEVSSARQVLOGPSATVNSPRENALSPEOLENLSAIKEYARDENEIG
 MCKDMSPAYHPTMISLAWTDSPEKDEKEDLAKLVNLSADNALMEVOLKGE
 ESVLKTELDADVNDAPKAEFLRFGSVTEKVVTLTEICRLIOEGGEESGLIEFL
 GGVLYKTELEMKTEACEEPTLVEIRSSGSENGETRGASIRKQIKMEVSOHSKYF
 CEFCGKTGVKRAKANGTGCCKDCGVKAGAYTMTASAVTVRSTIRLRQIGS"
 39540..39540
 /gene="F27H5_30"
 /number=1
 39541..42284
 /gene="F27H5_30"
 /number=2
 42285..42683
 /gene="F27H5_30"
 /number=2
 42684..42878
 /gene="F27H5_30"
 /number=3
 42879..42964
 /gene="F27H5_30"
 /number=3
 42965..43227
 /gene="F27H5_30"
 /number=4
 43228..44008
 /gene="F27H5_30"
 /number=4
 44009..44026
 /gene="F27H5_30"
 /number=5
 44027..44320
 /gene="F27H5_30"
 /number=5
 44321..44409
 /gene="F27H5_30"
 /number=6
 44410..44507
 /gene="F27H5_30"
 /number=6
 44508..44590
 /gene="F27H5_30"
 /number=7
 44591..44686
 /gene="F27H5_30"
 /number=7
 44687..44750
 /gene="F27H5_30"
 /number=8
 45114..45962
 /gene="F27H5_40"
 complement(join(45714..45844,45993..46158,46255..46364,
 46453..46671,46758..46962))
 /gene="F27H5_40"
 /codon_start=1
 /product="regulatory subunit of protein kinase CK2"
 /protein_id="CAB87862.1"
 /db_xref="GI:7576201"
 /translation="MKERSGGGGSSSFTLIGALIDKKRIDNALNKLKESSTSTT
 TSVFSSKDDPFSFTSKTQLPDVESEDTESGSDVSGSGDDTSTWISMCNLKGNDF

Query Match

74.4%; Score 18.6; DB 8; Length 88548;

Best Local Similarity 84.0%; Pred. No. 6, 4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 ggcagcagtgatcgccagcatt 25
 1 ||||| ||||| |||||
 Db 74472 GAGCAGCAGCTGCATTCCAGCATT 74448

RESULT 43
 ATT8B10/c
 LOCUS
 DEFINITION
 Arabidopsis thaliana DNA chromosome 3, BAC clone T8B10.
 AL138646
 VERSION
 AL138646.2 GI:7287982
 KEYWORDS
 SOURCE
 ORGANISM

thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 103787)
 Rieger, M., Mueller-Auer, S., Zipp, M., Schaefer, M., Mewes, H. W.,
 Lemcke, K., Mayer, K. F. X., Queletier, F. and Salanoubat, M.

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (20-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
 Coordinator: Marcel Salanoubat and Francis Queletier, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr

COMMENT
 On Mar 22, 2000 this sequence version replaced gi:689957.
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
 location/Qualifiers

FEATURES
 SOURCE
 1..103787
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 complement(8353..8556)
 /number=1
 8353..12143
 /gene="T8B10_10"
 complement(join(8353..8556,8641..8947,9034..9209,
 9288..9351,9485..9579,9673..9888,10013..10144,
 10459..10539,10632..12143))
 /note="Similarity to arm repeat containing protein
 homolog, Arabidopsis thaliana, EMBL:AL133314"
 homolog, Arabidopsis thaliana, EMBL:AL133314"
 /codon_start=1
 /product="Arm repeat containing protein-like"
 /protein_id="CAB81821.1"
 /db_xref="GI:7287983"
 /translation="MSRRVROVRVDNCKKVKDPSYTVIGVEDLAPKVVQYVYNTSLP
 VDPVPHLPFRNLNDRBRASLSTCRWMSLGSSTLSLNDPDLRAKPLDLSMAASLSTGC
 VDLOKIRFRGVSDADAIHLKARSLLETSCGYCKRTIDATLSMTAAHRLAESLQIP
 DECEKRTSDAIRVIAFCPCPKLKRVSQMRVSSDAIESLAKRHPQSLDGLFDCINL
 NEALGRVNSLRYSVAGTSNIKKRVALENNEKLPKLGIVSRFTTIDHIAVSRULKS
 SOSLKVLCALCPYLEEDKSYSNRFGKVLAVPTTFDELAIFADNSKPKNIFS
 YWRDLIRKDSIDEIMLWEMIISHTLIRLIESNSOGLNDFNLNOCATLILSLMOSEA
 OEDVOERATGATLAFIYVDENASIDCGRAVVRDGIIRLLELAKSRMGLOSSEA
 KAIANLSNAKAVARVAEGEISVIADLAKSMNLIVAEAGAGLMTNLSVEPERNATA
 QAGVNAVLADLIFRPHGCDGLERAKALNLAADKCSMEVARAGGVHVALNAN
 KYEGAQOAAALANLANAHDSNGNNAVQDEAGALEALVQLTOSFHEGYKORAA
 LMNLAFDDKRNRESIAAFGVEALVALAKSSNASTGLQERYAGALWGLSVSEANSIAT
 GHEGCIPLIALVREAEVDVETAGALAKNLSPNGNLRIVEGCVVALVQLCSSSV
 SKMARFMAALALVMEGRDMEYVIMIGLSLSTSKSVTLNCAFTMALDOIKAFTKTM
 EHOITFETASAPMSLMOVSERRAIPAGHLRCSGIGREVTMLRNPCLVLRACAA
 FALIQFTIPESHAMHHSIMONGEANGLSAAAAAAMPREAIKPKIYLRNHEHO

```

intron      complement(8557..8640)
exon        /number=1
intron      complement(8641..8947)
exon        /number=2
intron      complement(8948..9033)
exon        /number=2
intron      complement(9034..9209)
exon        /number=3
intron      complement(9210..9287)
exon        /number=3
intron      complement(9288..9351)
exon        /number=4
intron      complement(9352..9484)
exon        /number=4
intron      complement(9485..9579)
exon        /number=5
intron      complement(9580..9672)
exon        /number=5
intron      complement(9673..9888)
exon        /number=6
intron      complement(9889..10012)
exon        /number=6
intron      complement(10013..10144)
exon        /number=7
intron      complement(10145..10458)
exon        /number=7
intron      complement(10459..10539)
exon        /number=8
intron      complement(10540..10631)
exon        /number=8
intron      complement(10632..12143)
exon        /number=9
gene        /gene="T8B10.20"
CDS         complement(join(14024..14116,14198..14297,14381..14461,
14652..14731,14825..14918,15040..15131,15318..15408,
15494..15549))
            /note="similarity to CGF-94 protein, Homo sapiens,
            EMBL:AF151852"
            /codon_start=1
            /product="putative protein"
            /protein_id="CAB81822.1"
            /db_xref="GI:7287984"
            /translation="MSSLRNAIPRPAHKERSQPEARKRFGILEKHKDYIIIRANAAYHK
            QETLKITLRQAAAFKNPDEFNFKMINSTVTDGRHPKDEVAKYSAEELMIKTDIDIV
            FOKWQSEKKNIDKLTASLOCTGDQSSRRHYVAEDDEARELEVGGRSSTIETVP
            KDJKKMDRSYRDELEGRKSRKADLEKLYTMSMQKELQKKGRKRKLRLDELLNPNKA
            VYWRADRRK"
exon        complement(14024..14116)
intron      /number=1
intron      complement(14117..14197)
exon        /number=1
intron      complement(14198..14297)
exon        /number=2
intron      complement(14298..14380)
exon        /number=2
intron      complement(14381..14461)
exon        /number=3
intron      complement(14462..14511)
exon        /number=3
intron      complement(14512..14731)
exon        /number=4
intron      complement(14732..14824)
exon        /number=4
intron      complement(14825..14918)
exon        /number=5
intron      complement(14919..15039)
exon        /number=5
intron      complement(15040..15131)
exon        /number=6
intron      complement(15132..15317)
            /number=6

exon        complement(15318..15408)
intron      /number=7
exon        complement(15409..15493)
intron      /number=7
exon        complement(15494..15549)
intron      /number=8
gene        /number=8
CDS         /gene="T8B10.30"
            complement(join(16547..16646,17116..17228,17358..17467,
            17581..17851))
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CAB81823.1"
            /db_xref="GI:7287985"
            /translation="MYKFAILIRNPFSPFMPDSCDKAWSEFNSLSRSLVYLVASPC
            ILLPALSSAKTKSKSPYDERRLLEONKTIQENNAAPDEFPVEGEVAVASDNY
            IKNDLSLIYRDENVGOGDFPKDGOQVTFHYIGYNESGRIDSTYIGSPARIMQCTNA
            LVPENSKQFEVEDVELSLQNCERRITIGESDYVCS"
exon        /number=1
intron      complement(16547..16646)
intron      /number=1
intron      complement(16647..17115)
exon        /number=1
intron      complement(17116..17228)
exon        /number=2
intron      complement(17229..17357)
exon        /number=2
intron      complement(17358..17467)
exon        /number=3
intron      complement(17468..17580)
exon        /number=3
intron      complement(17581..17851)
exon        /number=4
gene        /number=4
CDS         /gene="T8B10.40"
            complement(18460..20691)
            /note="Contains ATP-dependent DNA ligase signatures
            AA577-585"
            /codon_start=1
            /product="putative protein"
            /protein_id="CAB81824.1"
            /db_xref="GI:7287986"
            /translation="MASPNPYTKRRSPNVVPPPOPKYKISGGGFECKSVLAFLL
            ALPLFSOAPDPEVGETVLTKPMELIHLIFGIAVAXGLFSRRNESAYDLMTVDSS
            SLSTVSRIFQVSVRDEEDFNDSCEFDVYRDESVARASVAGSEFVVEGLES
            SERGTENVRMANNSYFEGCKKVVAFPAIGLDGVVHQPLGPIRLRLBSLRLNAL
            IDKSFADSCDAVNAEASILLADNFEDEVLAAPSPVPMQARPPMMQIGDNYSPFP
            ISVDETAKTSSRSTSSQTSYASQNRSPSPSVSAESLNSNVEELVKEKRSOS
            SSRSSPSLPPSPSLSPSPPELVPNQTRRSPDELVTDPTRASHRSRYSDGSLSE
            EDVRGFEENELESGSVGRKAEPFSKKEGSKSLNLAESSRGKRRSYPPSIS
            PVGAGDSTTRRODLOOKSNCHLEENIRKGVADHNNLRVKKGRSHDSLPIAEDSA
            KDEKYESSEFPALDVVFOPTNAKASRRAMRSRGDRNLPKEDVYTRKLEDDTVDSKS
            RKDLPCKDKEMKLDGPRITPRSRNASSNVSIRKSVTIRSDHGVKDVTDSDSD
            RAEAKVBSRGCTKSRPQDELSTLVLDQESSFETRAKSEPEVAEMEPQAEQPEVTF
            EEEEAWEQSOSNASHDNEVDRKAGFEIARFROIQLQLISGEQPRGGGTGIFRNS
            QFR"
exon        complement(18460..20691)
intron      /number=1
intron      complement(22335..23917)
CDS         /gene="T8B10.50"
            complement(join(22335..22644,23038..23117,23202..23560,
            23719..23917))
            /note="Contains 'Homeobox' domain signature and profile
            AA192-215"
            /codon_start=1

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 8; Length 103787;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ggcagcaggtgcacgcagcagcatt 25
DB 1019 GAGCAGCAGCTGCATTTCACACATT 995

```



```

RESULT 44
ATT12K4/c 108879 bp DNA PLN 05-APR-2000
LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T12K4.
DEFINITION AL138640
ACCESSION AL138640
VERSION AL138640.1 GI:6899910
KEYWORDS
SOURCE
ORGANISM
thaie cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 108879)
Monfort,A., Casacuberta,E., Puigdememch,P., Mexes,H.W., Rued,S.,
Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.
Unpublished
2 (bases 1 to 108879)
EU Arabidopsis sequencing project.
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (05-APR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
source
Location/Qualifiers
1..108879
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
1..5515
/note="overlap to BAC T32A11, please refer also to
EMBL:AL138653 for analysis and annotation"
1099..5822
/note="copla-1-like LTR retrotransposon"
1099..1395
/note="297 bp long terminal repeat"
5526..5822
/note="297 bp long terminal repeat"
6818..7357
/note="28 bp long motif tgatttaccgttttgcggaataaata with
up to 9 mismatches repeated"
9190..9239
/note="50 bp at tandem repeat"
10223..10372
/number=1
/gene="T12K4_10"
10223..10718
/number=1
/gene="T12K4_10"
Join(10223..10372,10433..10504,10644..10718)
/gene="T12K4_10"
/note="similarity to crpi protein - Zea mays, PIR:T01685"
/codon_start=1
/product="putative protein"
/protein_id="CAB86439.1"
/db_xref="GI:7523381"
/translational="MENEGIEPNIIMSNVLYNMGAGTGWKLSPILILAKLYVYL
RAIKLDGTPDVVYITLMAKFTKPKRYEKNGSFWYGSSEKTAIAKCFYS"
10373..10432
/number=1
/gene="T12K4_10"
10433..10504
/number=1
/gene="T12K4_10"
10505..10643
/number=2
/gene="T12K4_10"
/number=2
exon
intron
10644..10718
/number=3
/gene="T12K4_10"
19016..20305
/number=1
/gene="T12K4_20"
Join(19016..19120,19958..20305)
/number=2
/gene="T12K4_20"
/note="similarity to peroxidase - Lycopersicon esculentum,
PIR:S32768"
/codon_start=1
/product="putative protein"
/protein_id="CAB86440.1"
/db_xref="GI:7523382"
/translational="METKKEKSGELIAKELKRLQTNVLRDLRFIDSHPSGDOQFTE
LDSEVNGIKKRLIGSLKTSLEHCEPKOVSCSVITLSARDYVALTGPLLISVLLGR
KDSLSTSKRHADSEPPSTADVDTLSTLASNOMTIGSVAILMGITL"
19016..19120
/number=1
/gene="T12K4_20"
19121..19957
/number=1
/gene="T12K4_20"
19958..20305
/number=1
/gene="T12K4_20"
20779..20814
/number=2
/gene="T12K4_30"
20779..25885
/number=1
/gene="T12K4_30"
Join(20779..20814,21087..21444,21675..21970,22038..22173,
22321..22451,22560..22679,22767..22793,22894..22998,
23085..23195,23291..23408,23487..23572,23659..23820,
23917..23994,24076..24148,24412..24470,24617..24928,
25268..25468,25583..25885)
/number=1
/gene="T12K4_30"
/note="similarity to hypothetical proteins - Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB86441.1"
/db_xref="GI:7523383"
/translational="MYHRRRLQWVRLPALDHYPTKAKINAYSKEIYIGITPAKNG
TPENTFLESPPGDLFRIPANKASFGKLVGLICRQVMKCYEMWVFGGHPRLG
LREWARLIGEGCTGPKNSAIEAVMOQEGRTSFLTDRCASFEGETDIEKLNCG
KOOSALHGFPLGIDLFAPETTPRSIASIVPEKQNTPTERSIKMLAKRLIKSTIL
ACDAKACTVAVVDVTVIVDYSQCPKSLVWRKCKQDPKPDNITLIMNLATSGMRK
ILKRRRVDPIDKRRKITNSDGGGSHSGGSGDSEDFDYADONEGRTASQKSDS
PSEKEMVNRKVKYKGNKRVHGVSRQAEHISLPTEILSSSDNVEDPNSQKGSFSDDE
KDIAPIOPVRNVPNGDGEVSARKHTTSDYVENIYLGASPHQEGNSIRE
DSSKIHONVSDSVMEEKFDSVPSPHLVTCIPDRDINEGLAOSKPIMSDVPSTHQ
DYAKDVNGTESVYAAELPKMEKSMSPKSPONPRVDEIDVAHEDKKRDEPT
SGDGEVLEKEAOSPSTVVEGVDTADQVPIVPPDTILAPFDEANLVNVEKESPIP
EKDYVEHPKRDVDVHMGVQSOTEEVNAHDEKERNVRRKVTAFADILKSGTISDHT
KSDVDFMETRESIINSVLVKAEDVPYNPLEKVDKAEKLSIMKEKMYALAKNG
EMKRRKRYVVDVPLWELSRATAFKSCVNEPFLDKLPLQFATGRLSNRERLRN
INDVDSVRSOTILPILEMPLPVATKCDYELKPLIIPFTYIRQRLSQNPPTGDCS
LYAMNPIELYMOMNNDLILIEANMNYNRGYAVDLNENHIGIST"
20815..21086
/number=1
/gene="T12K4_30"
21087..21444
/number=1
/gene="T12K4_30"
21445..21674
/number=2
/gene="T12K4_30"
21675..21970
/number=2
/gene="T12K4_30"
21971..22037
/number=3
/gene="T12K4_30"
/number=3
exon
intron

```

```

exon      22038..22173
          /gene="T12K4_30"
          /number=4
intron    22174..22320
          /gene="T12K4_30"
          /number=4
exon      22321..22451
          /gene="T12K4_30"
          /number=5
intron    22452..22559
          /gene="T12K4_30"
          /number=5
exon      22560..22679
          /gene="T12K4_30"
          /number=6
intron    22680..22766
          /gene="T12K4_30"
          /number=6
exon      22767..22793
          /gene="T12K4_30"
          /number=7
intron    22794..22893
          /gene="T12K4_30"
          /number=7
exon      22894..22998
          /gene="T12K4_30"
          /number=8
intron    22999..23084
          /gene="T12K4_30"
          /number=8
exon      23085..23195
          /gene="T12K4_30"
          /number=9
intron    23196..23290
          /gene="T12K4_30"
          /number=9
exon      23291..23408
          /gene="T12K4_30"
          /number=10
intron    23409..23486
          /gene="T12K4_30"
          /number=10
exon      23487..23572
          /gene="T12K4_30"
          /number=11
intron    23573..23658
          /gene="T12K4_30"
          /number=11
exon      23659..23820
          /gene="T12K4_30"
          /number=12
intron    23821..23916
          /gene="T12K4_30"
          /number=12
exon      23917..23994
          /gene="T12K4_30"
          /number=13

Query Match      74.4%: Score 18.6; DB 8; Length 108879;
Best Local Similarity 84.0%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgacgagtgcatgcgcagcatt 25
    ||||| ||||| ||||| |||||
Db 54730 GAGCAGCAGCGCATGCGCAGCATT 54706

RESULT 45
AC014784 120181 bp DNA HTG 16-NOV-1999
LOCUS AC014784
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014784

```

```

VERSION AC014784.1 GI:6436551
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 120181)
          Adams, M. and Venter, J.C.
          AUTHORS Direct Submission
          JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
          Rockville, MD, USA
COMMENT This sequence was identified as CDL:10209834 by the submitter.
          For further information on this sequence e-mail to fly@celera.com.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES
          location/Qualifiers
          1..120181
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
BASE COUNT 33873 a 26853 c 26669 g 32786 t
ORIGIN

```

```

Query Match      74.4%: Score 18.6; DB 2; Length 120181;
Best Local Similarity 84.0%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgacgagtgcatgcgcagcatt 25
    ||||| ||||| ||||| |||||
Db 32083 GCGGATCAGCTGCATCGCCAGCAGT 32059

Search completed: December 26, 2001, 10:56:58
Job time: 1980 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:18:29 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25

Sequence: 1 gcgcagcaggtgcagtcgcagcatt 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : N_Geneseq_1101.*

- 1: /SIDs2/gcgcdata/geneseq/geneseqn/NA1980.DAT.*
- 2: /SIDs2/gcgcdata/geneseq/geneseqn/NA1981.DAT.*
- 3: /SIDs2/gcgcdata/geneseq/geneseqn/NA1982.DAT.*
- 4: /SIDs2/gcgcdata/geneseq/geneseqn/NA1983.DAT.*
- 5: /SIDs2/gcgcdata/geneseq/geneseqn/NA1984.DAT.*
- 6: /SIDs2/gcgcdata/geneseq/geneseqn/NA1985.DAT.*
- 7: /SIDs2/gcgcdata/geneseq/geneseqn/NA1986.DAT.*
- 8: /SIDs2/gcgcdata/geneseq/geneseqn/NA1987.DAT.*
- 9: /SIDs2/gcgcdata/geneseq/geneseqn/NA1988.DAT.*
- 10: /SIDs2/gcgcdata/geneseq/geneseqn/NA1989.DAT.*
- 11: /SIDs2/gcgcdata/geneseq/geneseqn/NA1990.DAT.*
- 12: /SIDs2/gcgcdata/geneseq/geneseqn/NA1991.DAT.*
- 13: /SIDs2/gcgcdata/geneseq/geneseqn/NA1992.DAT.*
- 14: /SIDs2/gcgcdata/geneseq/geneseqn/NA1993.DAT.*
- 15: /SIDs2/gcgcdata/geneseq/geneseqn/NA1994.DAT.*
- 16: /SIDs2/gcgcdata/geneseq/geneseqn/NA1995.DAT.*
- 17: /SIDs2/gcgcdata/geneseq/geneseqn/NA1996.DAT.*
- 18: /SIDs2/gcgcdata/geneseq/geneseqn/NA1997.DAT.*
- 19: /SIDs2/gcgcdata/geneseq/geneseqn/NA1998.DAT.*
- 20: /SIDs2/gcgcdata/geneseq/geneseqn/NA1999.DAT.*
- 21: /SIDs2/gcgcdata/geneseq/geneseqn/NA2000.DAT.*
- 22: /SIDs2/gcgcdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA010403 E. coli biotin syn
3	25	100.0	1084	10	AA091329 E. coli B gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesi
6	18.4	73.6	860	21	AA015143 Trichoderma reesei
7	18.2	72.8	735	22	AAH03448 Human cDNA clone (
8	18.2	72.8	1746	22	AAH15264 Human cDNA sequenc
9	18.2	72.8	2531	20	AA211737 Human transport-as
10	18.2	72.8	2536	22	AA056388 ABC transport rela
11	18.2	72.8	2547	22	AAH15050 Human cDNA sequenc

C 12	18.2	72.8	109519	22	AA08693
C 13	17.8	71.2	300	21	AAA01491
C 14	17.8	71.2	1194	22	AA07922
C 15	17.8	71.2	1197	21	AAA64352
C 16	17.8	71.2	1197	21	AAA39343
C 17	17.8	71.2	1197	21	AAH49533
C 18	17.8	71.2	1197	22	AA088440
C 19	17.8	71.2	1197	22	AA07921
C 20	17.8	71.2	2048	17	AA026295
C 21	17.6	70.4	40	22	AA031064
C 22	17.6	70.4	384	21	AA03291
C 23	17.6	70.4	1215	21	AA060774
C 24	17.6	70.4	1344	21	AA256381
C 25	17.6	70.4	2126	19	AA041999
C 26	17.6	70.4	2126	20	AA211174
C 27	17.6	70.4	2214	22	AA061051
C 28	17.6	70.4	47961	22	AA030757
C 29	17.2	68.8	423	18	AA080740
C 30	17.2	68.8	568	21	AA098857
C 31	17.2	68.8	769	21	AA058894
C 32	17.2	68.8	997	22	AA118617
C 33	17.2	68.8	997	22	AA143731
C 34	17.2	68.8	999	22	AAH19796
C 35	17.2	68.8	1452	13	AA031876
C 36	17.2	68.8	1452	14	AA053208
C 37	17.2	68.8	1815	18	AA064495
C 38	17.2	68.8	1875	22	AA068252
C 39	17.2	68.8	1896	22	AA056335
C 40	17.2	68.8	2148	22	AA056338
C 41	17.2	68.8	2265	22	AA065400
C 42	17.2	68.8	2850	20	AA024652
C 43	17.2	68.8	2856	21	AA065871
C 44	17.2	68.8	2893	22	AA098174
C 45	17.2	68.8	5532	21	AAA81538
C 46	17.2	68.8	309400	22	AA068554
C 47	17.2	68.8	349980	22	AA021611
C 48	17.2	68.8	349980	22	AA068525
C 49	17.2	68.8	837096	21	AA081469
C 50	17.2	68.0	468	21	AAA55225
C 51	17.2	68.0	514	21	AA049411
C 52	17.2	68.0	828	20	AA084449
C 53	17.2	68.0	828	20	AA027714
C 54	17.2	68.0	911	21	AA024479
C 55	17.2	68.0	952	20	AA084443
C 56	17.2	68.0	952	20	AA027708
C 57	17.2	68.0	2320	21	AA021879
C 58	17.2	68.0	2568	22	AA064980
C 59	17.2	68.0	2905	21	AA044710
C 60	17.2	68.0	4946	22	AA088214
C 61	17.2	68.0	349980	21	AAA55186
C 62	17.2	68.0	349980	22	AA064966
C 63	16.8	67.2	531	22	AA06167
C 64	16.8	67.2	548	22	AA066339
C 65	16.8	67.2	2513	22	AAH14414
C 66	16.8	67.2	2623	22	AA021275
C 67	16.8	67.2	2778	22	AA070663
C 68	16.8	67.2	3030	22	AA070682
C 69	16.8	67.2	9551	20	AA022301
C 70	16.8	67.2	20394	22	AA024892
C 71	16.8	67.2	349980	22	AA068527
C 72	16.8	67.2	534720	19	AA030458
C 73	16.8	67.2	536165	19	AA030459
C 74	16.6	66.4	68	20	AA020565
C 75	16.6	66.4	70	20	AA020584
C 76	16.6	66.4	200	19	AA012833
C 77	16.6	66.4	200	19	AA011758
C 78	16.6	66.4	330	20	AA098047
C 79	16.6	66.4	363	20	AA021513
C 80	16.6	66.4	363	21	AA065000
C 81	16.6	66.4	414	21	AA063898
C 82	16.6	66.4	423	18	AA080632
C 83	16.6	66.4	423	18	AA080634
C 84	16.6	66.4	484	17	AA013967

Micromonospora DNA
Human colon cancer
Human G-protein co
cDNA encoding a hu
Human G-protein-co
Human GTP-binding
Human G-protein co
Human G-protein co
Pseudomonas sp Typ
NF-kappa binding
Human secreted pro
Human SERR ligand
Escherichia coli f
Human GPR14 (G-pro
Human GPR14 coding
P. putida KT2440-a
Micromonospora meg
Type II topoisomer
Fusarium venenatum
Human tumour suppl
Probe #8550 for ge
Probe #12417 used
Rhodospirillum rubrum
Cyclin D2 pseudoge
Human cyclin D2 ps
Human glucose tran
C glutamicum codin
DNA encoding Chlam
DNA encoding Chlam
C glutamicum codin
Human lung tumor a
Human lung cancer-
Human late stage o
N. meningitidis pa
C glutamicum codin
Neisseria meningit
C glutamicum codin
N. meningitidis pa
C. symbiosum open
Arabidopsis thalia
Membrane bound por
Active membrane-po
Bovine c-KIT GST-c
Porcine stem cell
Porcine stem cell
Human breast and o
C glutamicum codin
Arabidopsis thalia
Human DNA helicase
Cenarchaeum symbio
C glutamicum codin
C glutamicum codin
Human cDNA clone (f
Human cDNA sequenc
Human secreted pro
Human secreted pro
cDNA encoding a hu
Pimaric acid biosynt
C glutamicum codin
Rhizobium species
Rhizobium species
CTLA-4 variable 11
Human biallelic po
Human biallelic po
Nucleotide sequenc
Human CTLA-4 scaff
cDNA encoding a hu
Type II topoisomer
Type II topoisomer
Elmeria gametocyte

```

C 85 16.6 66.4 484 20 AAX89271 DNA sequence of cl
86 16.6 66.4 496 21 AAA45839 Human secreted exp
87 16.6 66.4 506 22 AAA42475 Human secreted exp
C 88 16.6 66.4 512 22 AA127113 Probe #17046 for g
C 89 16.6 66.4 512 22 AA155955 Probe #24641 used
C 90 16.6 66.4 553 22 AAH13249 Human cDNA clone (
91 16.6 66.4 561 17 AAT07213 Human T cell spec
92 16.6 66.4 561 19 AA40620 Soluble CTLA4 muta
93 16.6 66.4 561 20 AA223188 CTLA4 receptor wit
94 16.6 66.4 561 20 AAX26412 Human CTLA recepto
95 16.6 66.4 561 20 AAX26411 Human CTLA recepto
C 96 16.6 66.4 575 22 AAH07405 Human cDNA clone (
C 97 16.6 66.4 575 22 AA117923 Probe #1856 for ge
C 98 16.6 66.4 635 14 AA142902 Probe #11588 used
C 99 16.6 66.4 635 14 AA035132 Human CTLA4 recept
100 16.6 66.4 636 15 AA070451 DNA encoding human

```

ALIGNMENTS

RESULT 1

AA162941 ID AA162941 standard; DNA: 839 BP.

AC AA162941;

DT 22-OCT-2001 (first entry)

DE Human genomic DNA SEQ ID NO 269.

XX Human; nootropic; neuroprotective; cytosolic; dermatological; virocidic;
 XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 XX antiparkinsonian; antisticking; antianemic; antiarthritic; cancer;
 XX antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
 XX antiallergic; antidiabetic; antituber; anticonvulsant; antifungal;
 XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; nephrotropic; gene therapy; vaccine;
 XX ds.

OS Homo sapiens.

PN WO200155449-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01346.

XX 31-JAN-2000; 2000US-0179065.
 XX 04-FEB-2000; 2000US-0180628.
 XX 19-MAY-2000; 2000US-020515.
 XX 07-JUL-2000; 2000US-0216880.
 XX 14-JUL-2000; 2000US-0218290.
 XX 14-AUG-2000; 2000US-0225447.
 XX 01-SEP-2000; 2000US-0229343.
 XX 06-SEP-2000; 2000US-0230437.
 XX 08-SEP-2000; 2000US-0231243.
 XX 25-SEP-2000; 2000US-0234997.
 XX 29-SEP-2000; 2000US-0236367.
 XX 13-OCT-2000; 2000US-0239337.
 XX 08-NOV-2000; 2000US-0246476.
 XX 08-NOV-2000; 2000US-0246477.
 XX 08-NOV-2000; 2000US-0246525.
 XX 08-NOV-2000; 2000US-0246526.
 XX 08-NOV-2000; 2000US-0246528.
 XX 17-NOV-2000; 2000US-0249210.
 XX 17-NOV-2000; 2000US-0249211.
 XX 17-NOV-2000; 2000US-0249214.
 XX 01-DEC-2000; 2000US-0249265.
 XX 01-DEC-2000; 2000US-0250160.
 XX 05-DEC-2000; 2000US-0250391.
 XX 05-DEC-2000; 2000US-0251030.
 XX 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX PI
 XX PI Rosen CA, Barash SC, Ruben SM.
 XX DR WPI: 2001-476225/51.
 XX DX
 XX PT Novel plasma membrane associated proteins useful for diagnosing,
 XX PT treating, preventing and/or prognosing disorders related to the
 XX PT proteins, including cancer, immune response and neuronal disorders
 XX PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
 XX CC The invention relates to novel genes (AA162752-AA162961) and proteins
 XX CC (AA162347-AA162415) useful for preventing, treating or ameliorating
 XX CC medical conditions e.g. by protein or gene therapy. The genes are
 XX CC isolated from a range of human tissues disclosed in the specification.
 XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 XX CC and parasitic infections.
 XX CC Note: The sequence data for this patent did not form part of the
 XX CC printed specification, but was obtained in electronic format directly
 XX CC from WPI at ftp.wipo.int/pub/published_pat_sequences.
 XX SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
 Best Local Similarity 100.0%; Pred. No. 0.15; Mismatches 0; Gaps 0;
 Matches 25; Conservative 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcgcagcatt 25
 Db 111 ggcgcagcaggtgcacgcgcagcatt 135

RESULT 2
 ID AAX01303 standard; DNA: 1041 BP.

XX AAX01303;

DT 12-APR-1999 (first entry)

DE E. coli biotin synthetase (BioB) coding sequence.

XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
 XX biotin synthase; biotin production; vitamin H; BioB; ss.

OS Escherichia coli.

PN US5869719-A.

XX 09-FEB-1999.

XX 30-APR-1997; 97US-0846338.

XX 30-APR-1997; 97US-0846338.

XX 08-MAR-1995; 95US-0401068.

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Paton DA;
 XX
 DR WPI: 1999-152902/13.
 DR P-PSDB: AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di- amino- pelargonid acid amino- transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40: 34pp: English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC di- amino- pelargonid acid (DAP) aminotransferase or biotin synthase. The
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP: 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatgcgcagcatt 25
 ||||||||||||||||||||||||||||
 DB 79 gcgcagcaggtgcatgcgcagcatt 103

RESULT 3
 AAN91329
 ID AAN91329 standard: DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /tag-a
 XX
 GB2216530-A.
 XX
 PN 11-OCT-1989.
 XX
 PD 17-MAR-1989; 89GB-0006210.
 XX
 PF 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3: page 33-4; 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 XX
 SQ Sequence 1084 BP: 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatgcgcagcatt 25
 ||||||||||||||||||||||||||||
 DB 102 gcgcagcaggtgcatgcgcagcatt 126

RESULT 4
 AAN60496
 ID AAN60496 standard: DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetase enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag-a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PE 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB: AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 PS Disclosure: Page 534: 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 CC
 SQ Sequence 1121 BP: 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatgcgcagcatt 25
 ||||||||||||||||||||||||||||
 DB 120 gcgcagcaggtgcatgcgcagcatt 144

RESULT 5
 AAO62386
 ID AAO62386 standard: DNA: 5872 BP.
 XX

```

AC  AA062386;
XX  16-NOV-1994 (first entry)
DT  Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX  DE
XX  Biotin: expression; enterobacteria; vitamin H; synthesis;
KW  plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW  promoter; plac; biotin synthase; KAPA synthase;
KW  8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTR synthase;
KW  dehydrobiotin synthase; DAPA synthase;
KW  S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW  seborrhoea; dermatitis; ds.
XX  OS
XX  Escherichia coli DSM498.
FH  Key
FH  Location/Qualifiers
FT  1..96
FT  /tag= a
FT  /function= "promoter ptac"
FT  /evidence= EXPERIMENTAL
FT  23..28
FT  /tag= b
FT  /standard_name= "promoter ptac"
FT  45..50
FT  /tag= c
FT  /evidence= EXPERIMENTAL
FT  105..109
FT  /standard_name= "promoter ptac"
FT  105..109
FT  /tag= d
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "bioB RBS no. 9"
FT  117..1157
FT  /tag= e
FT  /product= "biotin synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioB"
FT  /number= 1
FT  1141..1146
FT  /tag= f
FT  /standard_name= "bioF RBS"
FT  1154..2311
FT  /tag= g
FT  /EC_number= 2.3.1.47
FT  /product= "KAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioF"
FT  /number= 2
FT  /standard_name= "8-amino-7-oxononanoate synthase"
FT  2284..2288
FT  /tag= h
FT  /standard_name= "bioC RBS"
FT  2295..3050
FT  /tag= i
FT  /function= "involved in pimeloyl-CoA synthesis"
FT  /product= "protein"
FT  /gene= "bioC"
FT  /number= 3
FT  3030..3033
FT  /tag= j
FT  /standard_name= "bioD RBS"
FT  3043..3753
FT  /tag= k
FT  /EC_number= 6.3.3.3
FT  /product= "DTR synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioD15"
FT  /number= 4
FT  /standard_name= "dehydrobiotin synthase"
FT  3712..3750
FT  /tag= l
FT  /note= "bioD15 substitution"
FT  3742..3746
RBS

```

```

FT  /tag= m
FT  /standard_name= "bioA RBS"
FT  3750..5039
FT  /tag= n
FT  /EC_number= 2.6.1.62
FT  /product= "DAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioA"
FT  /number= 5
FT  /standard_name= "S-adenosyl-L-methionine; 8-amino-
FT  5088..5093
FT  /tag= o
FT  /standard_name= "ORF1 RBS"
FT  5098..5574
FT  CDS
FT  /tag= p
FT  /function= "unknown, involved in biotin synthesis"
FT  /product= "protein"
FT  /evidence= EXPERIMENTAL
FT  /gene= "ORF1"
FT  /number= 6
FT  5583..5644
FT  terminator
FT  /tag= q
FT  /standard_name= "rho-independent transcriptional
FT  5583..5605
FT  /tag= r
FT  stem_loop
FT  5583..5605
FT  /tag= r
FT  WO9408023-A.
FT  PD  14-APR-1994.
FT  XX  14-APR-1994.
FT  XX  01-OCT-1993; 93WO-EP02688.
FT  PF  01-OCT-1993; 93WO-EP02688.
FT  XX  02-OCT-1992; 92CH-0003124.
FT  PR  15-JUL-1993; 93CH-0002134.
FT  XX  (LONZA AG.
FT  PA  (LONZA AG.
FT  XX  Birch O, Brass J, Fuhrmann M, Shaw N;
FT  PI  WPI: 1994-135587/16.
FT  DR  P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT  XX  Biotechnological biotin prodn. using enterobacterial biotin-gene
FT  PF  - providing vitamin H in high yield
FT  PT  Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT  XX  The sequence is derived from plasmid pB030A-15/9 contg. the
FT  PS  bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT  CC  of biotin, arranged in a transcription unit. Microorganisms
FT  CC  contg. these DNA fragments or plasmids may be used in the prodn.
FT  CC  of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT  CC  loss of appetite and tiredness.
FT  XX  Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

```

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcgcagcagtcgacccagcatt 25
    |||||
Db 195 gcgcagcagtcgacccagcatt 219

```

```

RESULT 6
AAFI5143
ID AAFI5143 standard; cDNA: 860 BP.
XX
AC AAFI5143;

```

XX 13-MAR-2001 (first entry)
 XX Trichoderma reesei EST SEQ ID NO:7666.
 DE Trichoderma reesei EST SEQ ID NO:7666.
 XX Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 XX Trichoderma reesei.
 OS WO200056762-A2.
 XX 28-SEP-2000.
 PD 22-MAR-2000; 2000WO-US07781.
 XX 22-MAR-1999; 99US-0273623.
 PR (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO NORDISK AS.
 PA Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 PI WPI; 2000-594572/56.
 DR Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX Claim 89: Page 3094; 3161pp; English.
 PS The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC Aspergillus niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
 CC niger; AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX Sequence 860 BP; 216 A; 249 C; 229 G; 141 T; 25 other;
 SQ

Query Match 73.6%; Score 18.4; DB 21; Length 860;
 Best Local Similarity 95.0%; Pred. No. 84;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 gcagcaggtgcacgcagcagc 22
 ||||||||| |||||
 DB 402 gcagcaggtgcacgcagcagc 421

RESULT 7
 AAH03448

ID AAH03448 standard; cDNA; 735 BP.
 XX AAH03448:
 AC 26-JUN-2001 (first entry)
 XX Human cDNA clone (5'-primer) SEQ ID NO:283.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 OS EP1074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 DR Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX Claim 1: SEQ ID 283; 2537pp + CD ROM; English.
 PS The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 735 BP; 203 A; 189 C; 215 G; 125 T; 3 other;
 SQ

Query Match 72.8%; Score 18.2; DB 22; Length 735;
 Best Local Similarity 87.0%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gcgcacaggtgcacgcagcagc 23
 ||||||||| |||||||||
 DB 686 gcgcacaggtgcacgcagcagc 708

```

RESULT      8
AAH15264
ID AAH15264 standard; cDNA; 1746 BP.
XX
AC AAH15264;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13394.
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8: SEQ ID 13394; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 1746 BP; 404 A; 471 C; 515 G; 356 T; 0 other;
XX
XX Query Match 72.8%; Score 18.2; DB 22; Length 1746;
XX Best Local Similarity 87.0%; Pred. No. 1.1e+02;
XX Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX 1 ggcagcaggtatcgcacga 23
XX |||||||| ||||||||
XX Db 686 ggcagcagcagtcgcacga 708

```

```

RESULT      9
AA211737
ID AA211737 standard; cDNA; 2533 BP.
XX
XX AA211737;
XX
XX 02-NOV-1999 (first entry)
XX
DE Human transport-associated protein-7 (TRANP-7) cDNA.
XX
XX Transport-associated protein; TRANP; nuclear pore; nuclear transport;
XX vesicle trafficking; cancer; cystic fibrosis; multidrug resistance;
XX hypercholesterolaemia; diagnosis; treatment; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 36..2165
XX FT /*tag= a
XX FT /product= "Human transport-associated protein-7"
XX
XX WO9941373-A2.
XX
XX 19-AUG-1999.
XX
XX 05-FEB-1999; 99WO-0502527.
XX
XX 11-FEB-1998; 98US-0021764.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Au-Young J, Bandman O, Baughn MR, Corley NC, Guegler KJ;
XX Hillman JL, Lal P, Yue H;
XX
XX WPI: 1999-508646/42.
XX
XX P-PSDB; AAY31645.
XX
XX Human TRANP coding sequences, used to treat transport disorders and
XX cancer
XX
XX Claim 7: Page 82-83; 87pp; English.
XX
XX This sequence represents human transport-associated protein-7 (TRANP-7)
XX cDNA. The DNA sequence was first identified in a human colon tissue
XX cDNA library. The full-length cDNA was derived from a series of
XX overlapping and/or extended cDNA sequences and is a consensus.
XX TRANP-1 to 9 (AAY31639-Y31647) are a novel group of proteins with
XX chemical and structural homology that are involved in molecular
XX transport. Various disorders are associated with defects in the transport
XX of molecules, either intracellularly or to the extracellular
XX environment. Examples of such disorders include cystic fibrosis,
XX multidrug resistance, hypercholesterolaemia and certain forms of diabetes
XX mellitus. Defective nuclear transport may play a role in cancer. For
XX example, the BRCA1 protein, associated with familial breast cancer, is
XX normally imported into the nucleus via nuclear pore complexes, but is
XX aberrantly located in the cytoplasm in breast cancer cells. In other
XX cancers, cells can secrete excessive amounts of hormones e.g. cancers of
XX the adrenal medulla can secrete excessive amounts of adrenaline and
XX noradrenaline, leading to hypertension. TRANP is expressed in cancer
XX cells, and transport disorders result from either excessive or
XX insufficient molecular transport. Anti-TRANP antibodies and nucleic
XX acids encoding TRANP can be used as diagnostic tools for such disorders. TRANP
XX antagonists can be used to treat or prevent a cancer associated with
XX increased TRANP expression. Anti-TRANP antibodies can be used directly
XX as an antagonist or as a targeting mechanism for drugs. Alternatively,
XX a TRANP antisense nucleotide can be used to treat cancers. A TRANP
XX agonist or expression vector may be used to treat a disorder caused by
XX reduced transport of biologically active molecules.
XX
XX Sequence 2533 BP; 610 A; 654 C; 756 G; 513 T; 0 other;

```


Query Match 72.8%; Score 18.2; DB 20; Length 2533;
 Best Local Similarity 87.0%; Pred. No. 1.1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatcgccagca 23
 ||||||||| | |||||||||
 Db 1320 gcgcagcagcagcatcgccagca 1342

RESULT 10

AA56388
 ID AAF56388 standard; DNA: 2536 BP.

XX AAF56388;

XX 23-APR-2001 (first entry)

XX ABC transport related gene #1.

XX Human; adenosine triphosphate-binding cassette;

XX transporter proteins; ABC; immune; infection; cancer; neurological;

XX cardiovascular; ds.

XX Homo sapiens.

XX WO200107658-A1.

XX 01-FEB-2001.

XX 20-JUL-2000; 2000WO-US19736.

XX 23-JUL-1999; 99US-0145215.

XX 18-AUG-1999; 99US-0149445.

XX 12-NOV-1999; 99US-0164730.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, NI J, Moore PE;

XX WPI: 2001-210950/21.

XX Claim 1; Page 262-263; 260pp; English.

XX The present invention relates to human adenosine

XX triphosphate-binding cassette (ABC) transporter proteins.

XX These proteins may be used to prevent, treat and diagnose

XX diseases associated with inappropriate expression of ABC transport

XX proteins. The disorders include for example: immune/autoimmune

XX diseases (e.g. HIV (human immunodeficiency virus) infections,

XX anemia, rheumatoid arthritis and multiple sclerosis), cancers

XX and hyperproliferative disorders, neurological diseases (e.g.

XX Alzheimer's disease, Parkinson's disease), cardio/cerebrovascular

XX disorders and infections caused by bacteria, viruses and fungi.

XX Sequence 2536 BP; 611 A; 655 C; 756 G; 514 T; 0 other;

Query Match 72.8%; Score 18.2; DB 22; Length 2536;

Best Local Similarity 87.0%; Pred. No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatcgccagca 23
 ||||||||| | |||||||||

Db 1323 gcgcagcagcagcatcgccagca 1345

RESULT 11

AAH15030
 ID AAH15030 standard; cDNA: 2547 BP.

XX AAH15030;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:13003.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Claim 8; SEQ ID 13003; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

XX in gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any special methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

XX represent oligonucleotides, all of which are used in the exemplification

XX of the present invention.

XX Sequence 2547 BP; 558 A; 673 C; 793 G; 523 T; 0 other;

OY 1 gcgcagcaggtgcatcgccagca 23
 ||||||||| | |||||||||

Db 1395 gcgcagcagcagcatcgccagca 1417

Query Match 72.8%; Score 18.2; DB 22; Length 2547;

Best Local Similarity 87.0%; Pred. No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
RESULT 12
AAS08693/C
ID AAS08693 standard; DNA; 109519 BP.
XX
AC AAS08693;
XX
DT 26-SEP-2001 (first entry)
XX
DE Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
KW Everninomycin; antibiotic; bottle-neck gene; orthonomycin;
XX
OS Micromonospora carbonacea var. africana.
XX
FH Key
FH Location/Qualifiers
FT complement (132..1382)
FT CDS
FT
FT
FT RBS
FT complement (1389..1394)
FT
FT CDS
FT complement (1490..2611)
FT
FT CDS
FT complement (2618..2622)
FT
FT CDS
FT complement (2622..3860)
FT
FT CDS
FT complement (3867..3870)
FT
FT CDS
FT complement (413..5312)
FT
FT CDS
FT complement (4134..4138)
FT
FT CDS
FT complement (5309..6235)
FT
FT CDS
FT complement (6232..7275)
FT
FT CDS
FT complement (6226..6229)
FT
FT CDS
FT complement (7272..8327)
FT
FT CDS
FT complement (8342..9364)
FT
FT CDS
FT complement (8333..8336)
FT
FT CDS
FT complement (9463..10224)
FT
FT CDS
FT complement (10232..10235)
FT
FT CDS
FT complement (10424..11176)
FT
FT CDS
FT complement (12027..12455)
FT
FT CDS
FT complement (12108..13022)
FT
FT CDS
FT complement (13027..13030)
FT
FT CDS
FT complement (14410..15363)
FT
```

```
FT
FT RBS
FT complement (15369..15373)
FT
FT CDS
FT complement (15380..16414)
FT
FT CDS
FT complement (16419..17873)
FT
FT CDS
FT complement (17870..18934)
FT
FT CDS
FT complement (19374..20906)
FT
FT CDS
FT complement (21064..22542)
FT
FT CDS
FT complement (22748..24172)
FT
FT CDS
FT complement (22736..22740)
FT
FT CDS
FT complement (24177..25223)
FT
FT CDS
FT complement (25230..25233)
FT
FT CDS
FT complement (25350..26626)
FT
FT CDS
FT complement (26685..30479)
FT
FT CDS
FT complement (26672..26676)
FT
FT CDS
FT complement (30557..31876)
FT
FT CDS
FT complement (31885..31888)
FT
FT CDS
FT complement (31941..32882)
FT
FT CDS
FT complement (33167..34405)
FT
FT CDS
FT complement (34414..34418)
FT
FT CDS
FT complement (34449..35210)
FT
FT CDS
FT complement (35219..35221)
FT
FT CDS
FT complement (35294..36238)
FT
FT CDS
FT complement (36235..36963)
FT
FT CDS
FT complement (36998..38026)
FT
FT CDS
FT complement (38072..38566)
FT
FT CDS
FT complement (38892..40163)
FT
FT CDS
FT complement (40216..40890)
FT
```

```

FT      /product= "EvrT"
FT      complement (40899..40902)
FT      /*lag= aw
FT      complement (40887..41576)
FT      /*lag= ax
FT      /product= "EvrU"
FT      complement (41679..42707)
FT      /*lag= ay
FT      /product= "EvrV"
FT      complement (42714..42717)
FT      /*lag= az
FT      complement (42810..43799)
FT      /*lag= ba
FT      /product= "EvrW"
FT      complement (43807..43811)
FT      /*lag= bd
FT      complement (43799..44866)
FT      /*lag= bc
FT      /product= "EvrX"
FT      complement (45014..45760)
FT      /*lag= bd
FT      /product= "EvrY"
FT      complement (45767..45770)
FT      /*lag= be
FT      /product= "EvrZ"
FT      complement (45952..45956)
FT      /*lag= bg
FT      complement (47156..49234)
FT      /*lag= bh
FT      /product= "EvsA"
FT      51627..52715
FT      /*lag= bi
FT      /product= "EvsB"
FT      51629..51622
FT      /*lag= bj
FT      52889..53557
FT      /*lag= bk
FT      /product= "EvsC"
FT      53554..54207
FT      /*lag= bl
FT      /product= "EvsD"
FT      complement (54362..55117)
FT      /*lag= bm
FT      /product= "EvsE"
FT      complement (55125..55128)
FT      /*lag= bn
FT      complement (55135..56094)
FT      /*lag= bo
FT      /product= "EvsF"
FT      complement (56100..56103)
FT      /*lag= bp
FT      complement (56184..56813)
FT      /*lag= bq
FT      /product= "EvsG"
FT      56961..58709

```

```

Query Match      72.8%; Score 18.2; DB 22; Length 109519;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY      1 gcgcagcagatgcacgcacga 23
      111111111111111111111111
Db      19913 GCGCGGCGAGTGCATCATCACGA 19891

```

```

RESULT 13
AAA01491/c
ID      AAA01491 standard; cDNA: 300 BP.
XX
AC      AAA01491;
XX

```

```

DT      19-MAY-2000 (first entry)
XX
XX      Human colon cancer cell line polynucleotide sequence SEQ ID NO:1482.
DE
XX      Human: colon cancer; tumour; diagnosis; gene expression product;
KW      Probe; detection; cancerous state; metastasis; identification;
KW      breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW      oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
XX      Homo sapiens.
XX
XX      W09958675-A2.
XX
XX      18-NOV-1999.
XX
XX      13-MAY-1999; 99WO-US10602.
XX
XX      14-MAY-1998; 98US-0085426.
XX      15-MAY-1998; 98US-0085537.
XX      15-MAY-1998; 98US-0085696.
XX      21-OCT-1998; 98US-0105234.
XX      27-OCT-1998; 98US-0105877.
XX
XX      (CHIR ) CHIRON CORP.
XX      (HYSE-) HYSEQ INC.
XX
XX      Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI      Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI      Lamsan G, Drmanac R, Crivenjakov R, Dickson M, Drmanac S, Labat I;
PI      Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX      WPI: 2000-126369/11.
XX
XX      Polynucleotide library used to determine cancerous states of mammalian
PT      cells -
PT
XX
XX      Claim 1; Page 536; 1097pp; English.
XX
XX      AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC      libraries constructed from human colon cancer cell lines. The present
CC      invention also describes a method of detecting differentially expressed
CC      genes correlated with a cancerous state of a mammalian cell, comprising
CC      detecting at least one differentially expressed gene product in a test
CC      sample derived from a cell suspected of being cancerous, where detection
CC      of the differentially expressed gene product is correlated with a
CC      cancerous state of the cell from which the test sample was derived.
CC      The polynucleotides sequences can be used in a method for detecting
CC      differentially expressed genes correlated with a cancerous state of a
CC      mammalian cell. The polynucleotides can also be used as probes for
CC      detecting and mapping related genes. They can be used in diagnosis and
CC      prognosis of diseases and disorders (e.g. identification of
CC      pre-metastatic or metastatic cancerous states, stages of cancer, or
CC      responsiveness of cancer to therapy). This is particularly for breast
CC      cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC      negative breast cancer, lung cancer, and colon cancer.
XX
XX      Sequence 300 BP; 56 A; 89 C; 94 G; 61 T; 0 other:
XX

```

```

Query Match      71.2%; Score 17.8; DB 21; Length 300;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      3 gcgcagcagatgcacgcacga 23
      111111111111111111111111
Db      108 GCAGCAGGTGCATCATCACGA 88

```

```

RESULT 14
AAD07922/c
ID      AAD07922 standard; cDNA: 1194 BP.
XX
AC      AAD07922;
XX

```


KM dysphagia; peptic oesophagitis; spasm; gastritis; anorexia; pyrosis;
 KM pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
 KM inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
 KM Addison's disease; allergy; asthma; diabetes mellitus; antihypoid;
 KM atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
 KM rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
 KM metabolic disorder; obesity; nootropic; prozoacide; vitruide; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 1..1194
 FT /tag= a
 FT /product= "Human GCRC-6 protein"
 XX
 PN WO200142288-A2.
 PD 14-JUN-2001.
 XX
 PF 07-DEC-2000; 2000WO-US33382.
 XX
 PR 10-DEC-1999; 990S-0172852.
 PR 22-DEC-1999; 990S-0171732.
 PR 14-JAN-2000; 2000US-0176148.
 PR 21-JAN-2000; 2000US-0177331.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Burford N, Baughn MR, Au-Young J, Yang J, Lu DM, Reddy R;
 XX
 DR WPI: 2001-381635/40.
 DR P-PSDB: AAE04550.
 XX
 PT New human G-protein coupled receptor polypeptides for diagnosing,
 PT preventing, and treating cell proliferative, neurological,
 PT cardiovascular, gastrointestinal, autoimmune and metabolic disorders -
 XX
 PS Claim 5; Page 162; 175pp; English.
 XX
 CC The present sequence is human G-protein coupled receptor-6 (GCRC-6)
 CC cDNA. GCRC is useful in somatic or germ-line gene therapy to correct a
 CC genetic deficiency, to express a conditionally lethal gene product and
 CC to express a protein which affords protection against intracellular
 CC parasites and also for diagnosis of disorders associated with expression
 CC of GCRC. GCRC is also useful for generating hybridisation probes useful
 CC in mapping the naturally occurring genomic sequences and to create
 CC knock-in humanised animals (pigs) or transgenic animals (mice or rats) to
 CC model human diseases. GCRC is used to diagnose, prevent and treat
 CC proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
 CC hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,
 CC uterus cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)
 CC neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,
 CC Parkinson's disease, multiple sclerosis, dementia and other central
 CC nervous system disorders); cardiovascular disorders (angina pectoris,
 CC hypertension, atherosclerosis, congestive heart failure);
 CC gastrointestinal disorders (dysphagia, peptic oesophagitis, oesophageal
 CC spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
 CC pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/
 CC inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
 CC Addison's disease, allergies, anaemia, asthma, diabetes mellitus, atopic
 CC dermatitis, glomerulonephritis, Grave's disease, osteoarthritis, atopic
 CC psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
 CC parasitic, protozoal and helminthic infections) and metabolic disorders
 CC (obesity, osteoporosis, viral infections).
 CC
 XX Sequence 1197 BP; 131 A; 482 C; 384 G; 200 T; 0 other;

DB 646 GCAGCAGTACGCCGACGCA 626
 ||||||||| || |||||||||
 RESULT 19
 AAD07921/c
 ID AAD07921 standard; cDNA: 1197 BP.
 XX
 AC AAD07921;
 XX
 DT 03-AUG-2001 (first entry)
 XX
 DE Human G-protein coupled receptor, AXOR12 cDNA.
 XX
 KW Human: G-protein coupled receptor; AXOR12; chromosome 19p13.3;
 KW microbial infection; human immunodeficiency virus; HIV; sleep disorder;
 KW pain; cancer; diabetes; obesity; anorexia; bulimia; asthma; gene therapy;
 KW Parkinson's disease; acute heart failure; hypotension; hypertension;
 KW urinary retention; osteoporosis; angina pectoris; myocardial infarction;
 KW stroke; ulcer; allergy; benign prostatic hypertrophy; migraine; anxiety;
 KW vomiting; dyskinesia; psychotic disorder; neurological disorder; vaccine;
 KW dementia; schizophrenia; depression; delirium; mental retardation; ss.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..1197
 FT /tag= a
 FT /product= "G-protein coupled receptor, AXOR12"
 XX
 PN WO200142486-A1.
 PD 14-JUN-2001.
 XX
 PF 08-DEC-2000; 2000WO-US33345.
 XX
 PR 08-DEC-1999; 990S-0456587.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Elishourbagy N, Shabon U, Michalovich D;
 XX
 DR WPI: 2001-381703/40.
 DR P-PSDB: AAE03544.
 XX
 PT New G-protein coupled receptor polypeptide, AXOR12, useful for treating
 PT microbial infections, cancers, obesity, asthma, diabetes, hypotension,
 PT osteoporosis, myocardial infarction and neurological disorders -
 XX
 PS Claim 2; Page 27; 37pp; English.
 XX
 CC The present sequence is human G-protein coupled receptor AXOR12 cDNA.
 CC The AXOR12 gene is located on human chromosome 19p13.3. AXOR12 protein
 CC and polynucleotide are useful for treating and diagnosing infections
 CC such as bacterial, fungal, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV)-1 or HIV-2,
 CC pain, cancers, diabetes, obesity, anorexia, bulimia, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, angina pectoris, myocardial
 CC infarction, stroke, ulcers, allergies, benign prostatic hypertrophy,
 CC migraine, vomiting, dyskinesias, sleep disorders, psychotic and
 CC neurological disorders including anxiety, schizophrenia, manic
 CC depression, depression, delirium, dementia and severe mental retardation.
 CC AXOR12 protein and polynucleotide are also useful for screening and
 CC structure based designing of antagonists, agonists and inhibitors of
 CC AXOR12. The protein is useful in screening assays and to
 CC identify membrane bound or soluble receptors. The polynucleotide is
 CC useful for chromosome localisation studies, as diagnostic reagents for
 CC detecting mutations in associated genes, recombinant production of AXOR12
 CC protein, as valuable tools for tissue expression studies and in gene
 CC therapy. AXOR12 protein and polynucleotide are also useful as vaccines.

QY 3 gcagcagtgatcgccagca 23
 Query Match 71.2%; Score 17.8; DB 22; Length 1197;
 Best Local Similarity 90.5%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Sequence 1197 BP; 131 A; 481 C; 385 G; 200 T; 0 other;

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 22; Length 1197;
Matches 19; Conservativity 90.5%; Pred. No. 1.5e+02;
Mismatches 2; Indels 0; Gaps 0;

OY 3 gcaagcagtgatgcagcagca 23
|||||
Db 646 GCAGCAGCTACGCCGACGCA 626

RESULT 20

AAAF26295/C
ID AAF26295 standard; DNA: 2048 BP.

XX AAF26295;

DT 02-MAY-2001 (first entry)

DE Pseudomonas sp Type II (xcp) secretion system DNA ORF04962a.

XX Type II secretion system; xcp secretion system; ORF04962a; ds.

XX Pseudomonas sp.

XX WO200107622-A2.

XX 01-FEB-2001.

XX 25-JUL-2000; 2000WO-EP07112.

XX 27-JUL-1999; 99DE-1035086.

XX (TIGR-) TIGR INST GENOMIC RES.

XX (GDB-) GDB GENOTYPE INFORMATION SYSTEM.

XX (GDB-) GDB GENOTYPE INFORMATION SYSTEM.

XX (DKFZ-) DKFZ DEUTSCHES KREBSFORSCHUNGSZENTRUM.

XX (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.

XX Fraser CM, Tuemmler B, Hoheisel J, Duesterhoeft A, Hilbert H;

XX Timmis KN, Moore E, Straetz M, Heim S, Nelson KE, Hickey E;

XX Peterson J;

XX WPI: 2001-159723/16.

XX New DNA encoding type II (xcp) secretion systems of the bacteria of

XX the genus Pseudomonas

XX Claim 3a; Page 24-25; 30pp; German.

XX This invention describes novel DNA sequences that encode type II (xcp)

XX secretion systems, especially type II (xcp) secretion systems of the

XX bacteria of the genus Pseudomonas, for example of the Pseudomonas putida

XX and Pseudomonas fluorescens group.

XX Sequence 2048 BP; 348 A; 712 C; 674 G; 314 T; 0 other;

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 22; Length 2048;
Matches 19; Conservativity 90.5%; Pred. No. 1.6e+02;
Mismatches 2; Indels 0; Gaps 0;

OY 3 gcaagcagtgatgcagcagca 23
|||||
Db 1596 GCAGCAGCTACGCCGACGCA 1576

RESULT 21
ID AAT31064/C
XX AAT31064 standard; CDNA: 40 BP.

AC AAT31064;

XX 06-FEB-1997 (first entry)

XX NF-kappaB binding site of Human Immunodeficiency Virus.

XX CD14; treatment: prevention; inflammation; sepsis; liver failure;

XX adult respiratory distress syndrome; acute pancreatitis;

XX inflammatory bowel disease; graft versus host disease; bone marrow;

XX transplant; tuberculosis; endotoxin contamination; detoxification;

XX ss.

XX Synthetic.

XX WO9620957-A1.

XX 11-JUL-1996.

XX 28-DEC-1995; 95WO-US17095.

XX 07-JUN-1995; 95US-0484397.

XX 30-DEC-1994; 94US-0366953.

XX (AMGE-) AMGEN INC.

XX (UYRQ) UNIV ROCKEFELLER.

XX Juan SC, Lichenstein HS, Wright SD;

XX WPI: 1996-333944/33.

XX Modified forms of CD14 polypeptide and DNA encoding them - useful

XX for treating and preventing inflammation, esp. sepsis

XX Example 8; Page 26; 79pp; English.

XX Derivatives of the CD14 polypeptide may be used to treat or prevent

XX inflammation, especially sepsis but also adult respiratory distress

XX syndrome, acute pancreatitis, acute or chronic liver failure,

XX inflammatory bowel disease, graft vs. host disease in bone marrow

XX transplant, tuberculosis etc. They may also be used for in vitro

XX detoxification and prevention of endotoxin contamination. See AAM00468-

XX 81. Whole cell extracts from U937 cells were prepared to assess

XX transcription factor NF-kappaB activation. Cells were seeded in

XX 6-well plates at a density of 1 million cells per well one day prior

XX to stimulation. For stimulation, purified soluble CD14 1-348,

XX soluble CD14 delta(57-64) or soluble CD14(7-10)A was added at a

XX final concentration of 20ng/ml with or without 20ng/ml of Re595 LPS

XX for 20 hours. Protein concentration of the whole cell extracts was

XX determined by micro BCA assay and ranged between 1.5-2

XX microg/microtitre. For examining the NF-kappaB complexes,

XX electrophoretic mobility shift assays were performed. This sequence

XX and its complement were annealed to generate a double stranded DNA

XX containing the NF-kappaB binding site of the HIV virus long terminal

XX repeat promoter. This annealed fragment was then filled with Klenow

XX fragment and radioactively labelled and used as a probe.

XX Sequence 40 BP; 6 A; 13 C; 14 G; 7 T; 0 other;

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 17; Length 40;
Matches 20; Conservativity 83.3%; Pred. No. 1.4e+02;
Mismatches 4; Indels 0; Gaps 0;

OY 1 gcaagcagtgatgcagcagcat 24
|||||
Db 25 GCGCTGACGCTGACGCCGCGCAT 2

RESULT 22

AC AAT31064/C
XX AAT31064 standard; CDNA: 384 BP.

AC AAT31064;

DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 3289.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Mline Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 DR P-PSDB: AAG03285.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 3289; 71pp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 CC
 XX
 SQ Sequence 384 BP; 86 A; 107 C; 130 G; 61 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 384;
 Best Local Similarity 83.3%; Pred. NO. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcagcagtgatcgcagcagcatt 25
 ||||| ||||| ||||| ||||| ||
 Db 181 cgcagcagtgatcgcagcagcagcatt 204

RESULT 23
 ID AAA60774/c
 XX AAA60774 standard; cDNA: 1215 BP.
 XX
 AC AAA60774;
 XX
 DT 27-OCT-2000 (first entry)
 XX
 DE Human SENR ligand nucleotide sequence SEQ ID NO:25.
 XX
 KW SENR; sensory epithelium neuropptide-like receptor; urolensin II;
 KW diagnosis; G protein-coupled receptor; hypertension; GPR14; hormone;
 KW kidney disease; regulator; central function; circulatory function;
 KW heart function; immune system function; digestive function;
 KW metabolic function; genital function; ss.
 XX

OS Homo sapiens.
 XX
 PN WO200032627-A1.
 XX
 PD 08-JUN-2000.
 XX
 PF 29-NOV-1999; 99WO-JP06649.
 XX
 PR 30-NOV-1998; 98JP-0338984.
 PR 04-FEB-1999; 99JP-0026848.
 PR 26-AUG-1999; 99JP-0239367.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Mori M, Abe M, Shimomura Y, Sugo T, Kitada C;
 XX
 DR WPI: 2000-412287/35.
 XX
 PT Urolensin peptides which are ligands for sensory epithelium
 PT neuropptide-like receptor (SENR) for diagnosis and treatment of
 PT hypertension -
 XX
 PS Example 20; Page 137; 147pp; Japanese.
 XX
 CC The present invention provides peptides which are ligands for sensory
 CC epithelium neuropptide-like receptor (SENR), and their amides, esters
 CC and salts. SENR is a G-protein coupled receptor protein (also known as
 CC GPR14), and the peptides which are ligands for it are forms of the
 CC peptide hormone urolensin II. The peptides can be used in the treatment
 CC and diagnosis of hypertension and kidney disease, and the development of
 CC drugs which are regulators of central functions, circulatory functions,
 CC heart functions, immune system functions, digestive functions, metabolic
 CC functions and genital functions. The present sequence represents a
 CC human SENR ligand nucleotide sequence, which is used in an example from
 CC the present invention.
 CC
 XX
 SQ Sequence 1215 BP; 141 A; 487 C; 393 G; 194 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 1215;
 Best Local Similarity 83.3%; Pred. NO. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcgcagcagtgatcgcagcagcatt 24
 ||||| ||||| ||||| ||||| ||
 Db 592 GCGCACGACGCCGATGCCGACGAT 569

RESULT 24
 ID AA256381/c
 XX AA256381 standard; DNA: 1344 BP.
 XX
 AC AA256381;
 XX
 DT 17-MAR-2000 (first entry)
 XX
 DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.
 XX
 KW Flagellin; fliC; antigen; detection; ds.
 KW Escherichia coli.
 OS
 PN WO9961458-A1.
 XX
 PD 02-DEC-1999.
 XX
 PF 21-MAY-1999; 99WO-AU00385.
 XX
 PR 21-MAY-1998; 98AU-0003634.
 XX
 PA (UNSY) UNIV SYDNEY.
 XX
 PI Reeves PR, Wang L;

XX WPI: 2000-072598/06.
 DR Novel nucleic acid molecule useful for the detection of flagellated
 XX bacterial strains in food, faeces, etc.
 PT
 PS Claim 3: Page 225; 245pp; English.
 XX
 CC AA256331 to AA256398 represent nucleic acid molecules (1) encoding all
 CC or part of an Escherichia coli flagellin protein except a protein
 CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
 CC invention also describes a method of detecting the presence of E. coli
 CC of a particular H serotype in a sample, comprising specifically
 CC hybridising a nucleic acid, preferably at least a pair, derived from a
 CC flagellating gene, specific for a particular flagellin gene associated
 CC with the H serotype, to any E. coli in the sample which contain the gene,
 CC and detecting any hybridised molecules, identifying the presence of that
 CC serotype in the sample. (1) are useful for: (1) detecting the presence
 CC of E. coli of H serotype in a sample by hybridising at least one or a
 CC pair of (1) to any E. coli in the sample and detecting the hybridised
 CC nucleic acid molecules; and (2) for detecting the presence of both O
 CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)
 CC to any E. coli present in the sample and detecting the hybridised
 CC nucleic acid molecules. (1) is particularly useful for detecting the
 CC combination of O and H antigen. Hybridised (1) when using at least one
 CC (1) is detected by southern blot analysis and, when using a pair of (1),
 CC is detected by polymerase chain reaction (PCR). AA256399 to AA256440
 CC represent primers used in the exemplification of the present invention.
 XX
 SQ Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 1344;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcagcagggtgcacgcacgacatt 25
 |||||
 Db 600 CGTAACCGGGCGCATGCCGACGATT 577

RESULT 25
 AA41999/c
 ID AAV41999 standard; DNA: 2126 BP.
 AC
 XX AAV41999:
 XX
 XX 20-NOV-1998 (first entry)
 DT
 XX
 DE Human GPR14 (G-protein coupled receptor polypeptide) gene.
 XX
 KW G-protein coupled receptor polypeptide; GPR14; human; fungal infection;
 KW bacterial infection; protozoan infection; viral infection; agonist;
 KW antagonist; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 694..1863
 FT /*tag= a
 FT /product= "human GPR14"
 XX
 XX EP859052-A1.
 XX
 PD 19-AUG-1998.
 XX
 PF 18-NOV-1997; 97EP-0309252.
 XX
 PR 27-JAN-1997; 97US-0789354.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 XX
 PI Bergsma DJ, Shabon U;
 XX

XX WPI: 1998-429680/37.
 DR P-PSDB; AAW59908.
 XX
 PT New DNA sequence encoding G-protein coupled receptor polypeptide
 PT GPR14 - and corresponding polypeptide, agonists, antagonists, etc.
 XX
 PS Claim 3: Fig 1; 23pp; English.
 XX
 CC This is the nucleotide sequence of the gene encoding the novel G-protein
 CC coupled receptor polypeptide (GPR14), used in the method of the
 CC invention. Human GPR14 polypeptides and polynucleotides can be used in
 CC the treatment of infections such as bacterial, fungal, protozoan and
 CC viral infections. Agonists and antagonists can be used to treat
 CC conditions associated with Human GPR14 imbalance.
 XX
 SQ Sequence 2126 BP; 317 A; 763 C; 669 G; 377 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 2126;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcgcagcagggtgcacgcacgacatt 24
 |||||
 Db 1269 CGCGACGACCGCATGCCGACGATT 1246

RESULT 26
 AA21174/c
 ID AA21174 standard; cDNA: 2126 BP.
 AC
 XX AA21174:
 XX
 DT 04-NOV-1999 (first entry)
 DE
 XX
 DE Human GPR14 coding sequence.
 XX
 KW GPR14; G protein-coupled receptor; ischaemic coronary artery disease;
 KW atherosclerosis; metabolic disease; CHF/myocardial dysfunction; migraine;
 KW arhythmia; restenosis; hypertension; pulmonary disease;
 KW fibrotic vasculopathy; cerebrovascular event; neurogenic inflammation;
 KW haematopoietic disorder; adult respiratory distress syndrome; ARDS;
 KW cancer; autoimmune disease; therapy; human; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 694..1863
 FT /*tag= a
 FT /product= GPR14
 XX
 XX WO9940192-A1.
 XX
 PD 12-AUG-1999.
 XX
 PF 27-JAN-1999; 99WO-US01634.
 XX
 PR 15-JAN-1999; 99US-0232857.
 PR 09-FEB-1998; 98US-0074075.
 PR 10-APR-1998; 98US-0058725.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM LAB PHARM.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Alvar NV, Al-Barazani K, Ames RS, Arnold AR, Bergsma DJ,
 PI Chambers J, Douglas SA, Foley JJ, Gout B, Khandoudi N;
 PI Sarau HM, Shabon U, Willette RN;
 XX
 DR WPI: 1999-527305/44.
 DR P-PSDB; AAY32920.
 XX

PT Human G protein-coupled receptor GPR14, useful for identifying
 PT agonists and antagonists
 XX
 PS Claim 4; Fig 1: 64pp; English.
 CC This sequence encodes the human G protein-coupled receptor, GPR14, of the
 CC invention. Human GPR14, polypeptide and polynucleotides are useful in
 CC methods for treatment of, e.g. ischaemic coronary artery disease,
 CC atherosclerosis, metabolic diseases, CHF/myocardial dysfunction,
 CC arrhythmias, restenosis, hypertension and hypotension, pulmonary disease,
 CC fibrotic vasculopathies, cerebrovascular events, neurogenic
 CC inflammation/migraine, haematopoietic disorders, adult respiratory
 CC distress syndrome (ARDS), Cancer, autoimmune diseases, etc. The methods
 CC can be used to identify agonists and antagonists of human and rat GPR14.
 CC
 SQ Sequence 2126 BP; 317 A; 763 C; 669 G; 377 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 20; Length 2126;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 0Y 1 gccgacgagtgatcgcacgacat 24
 Db 1269 GCGACACGACCGCATGCGCAGCAT 1246
 ||||| ||||| ||||| ||||| |||||
 RESULT 27
 AAF61051/C
 ID AAF61051 standard; DNA: 2214 BP.
 XX
 AC AAF61051;
 XX
 DT 16-MAY-2001 (first entry)
 XX
 DE P. putida KT2440-associated DNA ORF07225.
 XX
 DE Transgenic plant; detection: probe; amplification; vaccine carrier;
 XX
 KM microbial production strain; biological remediation; ds.
 XX
 OS Pseudomonas putida.
 XX
 PN DEL19935088-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 27-JUL-1999; 99DE-1035088.
 XX
 PR 27-JUL-1999; 99DE-1035088.
 XX
 PA (TIGR-) TIGR INST GENOMIC RES.
 PA (QUIA-) QUIAGEN GMBH.
 PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX
 DR WPI; 2001-192469/20.
 XX
 PT New DNA sequences specific for Pseudomonas putida KT2440, useful as
 PT safe genetic engineering host, allow detection in presence of other
 PT related bacteria -
 XX
 PS Claim 1a; Page 101-102; 158pp; German.
 CC This invention describes novel DNA sequences (1) for specific detection
 CC of Pseudomonas putida KT2440. The invention also describes (1)
 CC recombinant expression vector containing (1); (2) prokaryotic or
 CC eukaryotic cells transformed or transfected with (1) or the vector of
 CC (1); (3) production of expression products by culturing cells of (2);
 CC (4) expression products, or their fragments, of (1) and synthetic
 CC proteins or peptides with the same sequences (A); (5) poly- or
 CC monoclonal antibodies (Ab) that react specifically with (A); (6)
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic

CC plants that contain transformed or transfected cells of (2); (8)
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
 CC carrying one or more (1) (11), and their fragments, are used as probes
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants. (11),
 CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (11) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen P. aeruginosa). Compared with other "safe" bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.
 CC
 SQ Sequence 2214 BP; 436 A; 713 C; 693 G; 372 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 22; Length 2214;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 0Y 1 gccgacgagtgatcgcacgacat 24
 Db 783 GCGCTGCGAGTGCGACGCGCCCAT 760
 ||||| ||||| ||||| ||||| |||||
 RESULT 28
 AAF30757
 ID AAF30757 standard; DNA: 47981 BP.
 XX
 AC AAF30757;
 XX
 DT 21-JUN-2001 (first entry)
 XX
 DE Micromonospora megalomicea megalomicin biosynthetic gene cluster.
 XX
 DE Megalomicin; meg gene; polyketide synthase; antibiotic;
 XX
 KM molillide; antiparasitic; ds.
 XX
 OS Micromonospora megalomicea subsp. nigra.
 XX
 FH Location/Qualifiers
 FT CDS complement (1..144)
 FT /tag= a
 FT /partial
 FT /gene= "megBVI(megT)"
 FT /product= "TDP-4-keto-6-deoxyglucose-
 FT 2,3-dehydratase"
 FT /note= "encodes AAB82201"
 FT 928..2061
 FT /tag= b
 FT /gene= "megDVI"
 FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
 FT /note= "encodes AAB82202"
 FT 2072..3382
 FT /tag= c
 FT /gene= "megDI"
 FT /product= "TDP-megosamine glycosyltransferase"
 FT /note= "erycIII homologue; encodes AAB802203"
 FT 3462..4634
 FT /tag= d
 FT /gene= "megY"
 FT /product= "mycarose O-acyltransferase"
 FT /note= "encodes AAB82204"
 FT 4651..5775
 FT /tag= e
 FT /gene= "megDII"
 FT /product= "TDP-3-keto-6-deoxyhexose
 FT 3-aminotransaminase"
 FT /note= "erycI, DnrJ homologue, encodes AAB82205"
 FT 5822..6595
 FT CDS

```
FT      /*tag= f
FT      /gene= "megDIIf"
FT      /product= "daunosaminyl-N,N-dimethyltransferase"
FT      /note= "eryCVI homologue; encodes AAB82206"
FT      6592..7197
FT      /*tag= g
FT      /gene= "megDIV"
FT      /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT      /note= "eryBVIII, dnmV homologue, encodes AAB82207"
FT      7220..8206
FT      /*tag= h
FT      /gene= "megDV"
FT      /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT      /note= "eryVIV, dnmV homologue; encodes AAB82208"
FT      complement (8228..9220)
FT      /*tag= i
FT      /gene= "megBII-1(megDVII)"
FT      /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT      /note= "encodes AAB82209"
FT      complement (9226..10479)
FT      /*tag= j
FT      /gene= "megBV"
FT      /product= "TDP-mycarose glycosyltransferase"
FT      /note= "encodes AAB82210"
FT      complement (10483..11424)
FT      /*tag= k
FT      /gene= "megBIV"
FT      /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT      /note= "encodes AAB82211"
FT      12181..22821
FT      /gene= "megaA"
FT      /*tag= l
FT      /product= "megalomicin 6-deoxyerythronolide B
FT      synthase 1"
FT      /note= "polyketide synthase; encodes AAB82212"
FT      12505..13470
FT      /*tag= m
FT      /gene= "megaA"
FT      /function= "At-L"
FT      13576..13791
FT      /*tag= n
FT      /gene= "megaA"
FT      /function= "ACP-L"
FT      13849..15126
FT      /*tag= o
FT      /gene= "megaA"
FT      /function= "KSI"
FT      15427..16476
FT      /*tag= p
FT      /gene= "megaA"
FT      /function= "AT1"
FT      17155..17694
FT      /*tag= q
FT      /gene= "megaA"
FT      /function= "KRI"
FT      17947..18207
FT      /*tag= r
FT      /gene= "megaA"
FT      /function= "ACPI"
FT      18268..19548
FT      /*tag= s
FT      /gene= "megaA"
FT      /function= "KS2"
FT      19876..20910
FT      /*tag= t
FT      /gene= "megaA"
FT      /function= "AT2"
FT      21517..22053
FT      /*tag= u
FT      /gene= "megaA"
FT      /function= "KR2"
FT      22318..22575
FT      /*tag= v
FT      misc_feature

FT      /gene= "megaA"
FT      /function= "ACP2"
FT      22867..33555
FT      /*tag= w
FT      /gene= "megAII"
FT      /product= "megalomicin 6-deoxyerythronolide B
FT      synthase 2"
FT      /note= "polyketide synthase, encodes AAB82213"
FT      22957..24237
FT      /*tag= x
FT      /gene= "megAII"
FT      /function= "KS3"
FT      24544..25581
FT      /*tag= y
FT      /gene= "megAII"
FT      /function= "AT3"
FT      26230..26733
FT      /*tag= z
FT      /gene= "megAII"
FT      /function= "KR3 (inactive)"
FT      26998..27258
FT      /*tag= aa
FT      /gene= "megAII"
FT      /function= "ACP3"
FT      27393..28590
FT      /*tag= ab
FT      /gene= "megAII"
FT      /function= "KS4"
FT      28897..29931
FT      /*tag= ac
FT      /gene= "megAII"
FT      /function= "AT4"
FT      29953..30477
FT      /*tag= ad
FT      /gene= "megAII"
FT      /function= "DH4"
FT      31396..32244
FT      /*tag= ae
FT      /gene= "megAII"
FT      /function= "ER4"
FT      32257..32799
FT      /*tag= af
FT      /gene= "megAII"
FT      /function= "KR4"
FT      33052..33312
FT      /*tag= ag
FT      /gene= "megAII"
FT      /function= "ACP4"
FT      33666..43271
FT      /*tag= ah
FT      /gene= "megAII"
FT      /product= "megalomicin 6-deoxyerythronolide B
FT      synthase 3"
FT      /note= "polyketide synthase; encodes AAB82214"
FT      22957..24237
FT      /*tag= ai
FT      /gene= "megAII"
FT      /function= "KS5"
FT      24544..25581
FT      /*tag= aj
FT      /gene= "megAII"
FT      /function= "AT5"
FT      26230..26733
FT      /*tag= ak
FT      /gene= "megAII"
FT      /function= "KR5"
FT      26998..27258
FT      /*tag= al
FT      /gene= "megAII"
FT      /function= "ACP5"
FT      27393..28590
FT      /*tag= am
FT      /gene= "megAII"
FT      misc_feature
```

```

FT      misc_feature      /function= "KS6"
FT      2897..29931
FT      /tag= an
FT      /gene= "megAIII"
FT      /function= "AT6"
FT      29953..30477
FT      /tag= ao
FT      /gene= "megAIII"
FT      /function= "KR6"
FT      31396..32244
FT      /tag= ap

Query Match      70.4%; Score 17.6; DB 22; Length 47981;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 ggcagcaggtgcatgcgcagcat 24
        ||||| || ||||| |||
Db      15792 ggcagcgcgggcatgcgcacat 15815

RESULT 29
AAT80740/c
ID      AAT80740 standard; DNA: 423 BP.
XX
AC      AAT80740;
XX
DT      11-NOV-1997 (first entry)
XX
DE      Type II topoisomerase database reference sequence SEQ ID NO:159.
XX
KW      Detection; identification; TopoII; contamination; food; human;
KW      animal; plant; soil; water; PCR; polymerase chain reaction; ds.
XX
OS      Nocardia farcinica.
XX
XX      US5645994-A.
XX
PD      08-JUL-1997.
XX
PF      05-JUL-1990; 900S-0548138.
XX
PR      06-JUN-1995; 950S-0470179.
PR      05-JUL-1990; 900S-0548138.
PR      13-AUG-1993; 930S-0106482.
XX
PA      (UTAH ) UNIV UTAH RES FOUND.
XX
PI      Huang WM;
XX
DR      WPI: 1997-362925/33.
XX
PT      Detection and identification of organisms - using DNA primers to
PT      amplify signature segment of organism's type II topoisomerase
XX
PS      Claim 6; Column 157-158; 114pp; English.
XX
CC      A method has been produced for selectively amplifying DNA segments of
CC      one or more species of organisms in a sample. The method involves: (a)
CC      providing a database containing reference sequences, comprising a
CC      subunit sequence of a signature region of a macromolecule selected from
CC      a type II topoisomerase (TopoII), or a homologue, where each reference
CC      sequence is specific to a different species of a chosen group, and the
CC      macromolecule comprises 1st and 2nd conserved regions adjacently
CC      flanking the signature region; and (b) making an extract of DNA
CC      molecules, and selectively amplifying DNA segments of the signature
CC      region using a universal primer composition, comprising a primer
CC      constructed to bind a DNA encoding the macromolecule, to produce
CC      amplified DNA segments. The present sequence represents a DNA fragment
CC      containing a signature segment which is used in the database as a
CC      reference sequence for Nocardia farcinica. The method can be
CC      used to identify all of the different organisms present in a single
CC      sample without using multiple probes. It can accurately distinguish

```

```

CC      among similar and related species, and can be used with, e.g.
CC      contaminated food products, tissue or fluid samples from diseased
CC      humans, animal or plants, soil samples and water samples from any
CC      source.
XX
SQ      Sequence 423 BP; 84 A; 144 C; 135 G; 60 T; 0 other;

Query Match      68.8%; Score 17.2; DB 18; Length 423;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ggcagcaggtgcatgcgcagcat 22
        ||||| || ||||| |||
Db      294 GCGAGCATCTCATCCAGC 273

RESULT 30
AAF09857/c
ID      AAF09857 standard; cDNA: 568 BP.
XX
AC      AAF09857;
XX
DT      13-MAR-2001 (first entry)
XX
DE      Fusarium venenatum EST SEQ ID NO:2380.
XX
KW      Multiple gene expression; filamentous fungal cell; EST;
KW      expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW      Aspergillus oryzae; Trichoderma reesei; identification;
KW      culture condition; environmental stress; spore morphogenesis;
KW      metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX      Fusarium venenatum.
XX
XX      WO200056762-A2.
XX
PD      28-SEP-2000.
XX
PF      22-MAR-2000; 2000WO-US07781.
XX
PR      22-MAR-1999; 99US-0273623.
XX
PA      (NOVO ) NOVO NORDISK BIOTECH INC.
PA      (NOVO ) NOVO NORDISK AS.
XX
PI      Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX      WPI: 2000-594572/56.
XX
DR      WPI: 2000-594572/56.
XX
PT      Monitoring differential expression of genes in filamentous fungal cells
PT      uses fluorescence-labeled nucleic acids isolated from the cells and a
PT      substrate of expressed sequence tags -
XX
PS      Claim 86; Page 1260; 3161pp; English.
XX
CC      The present invention describes a method for monitoring differential
CC      expression of genes in a first filamentous fungal (FF) cell relative to
CC      expression of the same genes in one or more second filamentous fungal
CC      cells. The method uses fluorescence-labeled nucleic acids isolated from
CC      the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC      are used in the methods for monitoring differential expression of genes
CC      in a first filamentous fungal (FF) cell relative to expression of the
CC      same genes in one or more second filamentous fungal cells. Monitoring
CC      the global expression of genes from FF cells allows the production
CC      potential of the microorganisms to be improved. New genes may be
CC      discovered, possible functions of unknown open reading frames can be
CC      identified and gene copy number variation and stability can be
CC      monitored. The expression of genes can be used to study how FF cells
CC      adapt to changes in culture conditions, environmental stress, spore
CC      morphogenesis, recombination, metabolic or catabolic pathway
CC      engineering. Using ESTs provides several advantages over genomic or
CC      random cDNA clones including elimination of redundancy as one spot on an

```

CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
 CC Fusarium venenatum; AAF1248 to AAF1853 represents ESTs from Aspergillus
 CC niger; AAF1854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

XX Sequence 568 BP; 127 A; 155 C; 160 G; 114 T; 12 other;

Query Match 68.8%; Score 17.2; DB 21; Length 568;
 Best Local Similarity 86.4%; Pred. No. 2.6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 gcagcaggtgcacgcagcat 24
 ||||| ||||| ||||| |||||
 Db 151 GCAGGAGGTGCCTCGCACGCT 130

RESULT 31
 AAC58894/C
 ID AAC58894 standard; DNA; 769 BP.

XX AAC58894;

DT 25-JAN-2001 (first entry)

XX Human tumour suppressor BRG1 gene exon 24.

XX Human BRG1: tumour suppressor gene; cancer; chromosome 19p13.1;

KW retinoblastoma tumour suppressor gene; RB; drug screening; gene therapy;
 drug design; peptide therapy; animal model; ss.

XX Homo sapiens.

XX WO200056931-A1.

XX 28-SEP-2000.

XX 23-MAR-2000; 2000WO-US07678.

XX 23-MAR-1999; 99US-0125806.

XX (MYRI-1) MYRIAD GENETICS INC.

XX Wong AKC, Tavligian SV, Teng DH;

XX WPI; 2000-587668/55.

PT Diagnosing a polymorphism associated with predisposition for cancer in
 PT humans by determining whether there is a germline alteration of a BRG1
 PT gene or its expression products -

XX Claim 18; Page 106; 215pp; English.

CC The present invention is concerned with the use of the human tumour
 CC suppressor gene BRG1 in cancer diagnosis and therapy. This gene is
 CC comprised of several exons, shown in AAC58874-C58903, and has several
 CC splice variants, given in AAC58906-C58912. The protein sequences for
 CC these are shown in AAB27552-B27558. BRG1 is a homologue of the Drosophila
 CC protein brhma, and has been shown to be bound to retinoblastoma tumour
 CC suppressor protein RB. The BRG1 coding sequence and protein can be used
 CC in the diagnosis and treatment of cancer (for example by gene therapy),
 CC particularly prostate cancer, to identify drugs useful in the treatment
 CC of cancer and in the production of animal models for cancer.

XX Sequence 769 BP; 139 A; 212 C; 222 G; 196 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 769;
 Best Local Similarity 86.4%; Pred. No. 2.6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 gcagcaggtgcacgcagcat 24
 ||||| ||||| ||||| |||||
 Db 234 GCAGGAGGTGCATGCCGCGCT 213

RESULT 32
 AA118617
 ID AA118617 standard; DNA; 997 BP.

XX AA118617;

DT 12-OCT-2001 (first entry)

XX Probe #8550 for gene expression analysis in human cervical cell sample.

KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632365.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 8550; 487np; English.

XX The present invention relates to human single exon nucleic acid probes
 XX (SENPs). The present sequence is one such probe. The SENPs are derived
 XX from human HeLa cells. The SENPs can be used to produce a single exon
 XX microarray, which can be used for measuring human gene expression in a
 XX sample derived from human cervical epithelial cells. By measuring gene
 XX expression, the probes are therefore useful in grading and/or staging
 XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 997 BP; 210 A; 258 C; 320 G; 209 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 997;
 Best Local Similarity 86.4%; Pred. No. 2.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gcagcaggtgcacgcagca 23
 ||||| ||||| ||||| |||||
 Db 293 ccagcaggtgcacgcagca 314

RESULT 33
 AA143731
 ID AA143731 standard; DNA; 997 BP.

```

AC AA143731:
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #12417 used to measure gene expression in human placenta sample.
DE
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200157272-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0612366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DK;
PI WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 12417; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 997 BP; 210 A; 258 C; 320 G; 209 T; 0 other;
SQ

```

Query Match 68.8%; Score 17.2; DB 22; Length 997;
 Best Local Similarity 86.4%; Pred. No. 2.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 2 cgcagcaggtgcatgcagca 23
   | | | | | | | | | | | | | | |
Db 293 cccagcaggtcctcctccagca 314

RESULT 34
AAH19796/C
ID AAH19796 standard; DNA; 999 BP.
XX
XX AAH19796:
AC
XX 30-JUL-2001 (first entry)
XX
XX Rhodobacter sphaeroides decaprenyl diphosphate synthase DNA.
DE
XX Rhodobacter sphaeroides; ubiquinone-10; microorganism; DPPS; cardiant;
KM antioxidant; decaprenyl diphosphate synthase; geranylgeranyl transferase;
KM hydroxybenzoic acid-decaprenyl transferase; antioxidant; heart disease;
XX ds.
XX Rhodobacter sphaeroides.
OS
XX Key Location/Qualifiers
FH CDS 1..999
FT

```

```

FT /*tag= a
FT /partial
FT /product= "decaprenyl diphosphate synthase (DPPS)"
FT /note= "no stop codon given"
XX
XX WO200127286-A1.
PN
XX 19-APR-2001.
PD
XX 13-OCT-2000; 2000WO-JP07121.
XX
XX 14-OCT-1999; 99JP-0291959.
PR
XX (RYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Miyake K, Hashimoto S, Ozaki A;
PI WPI; 2001-282036/29.
XX
XX P-PDB; AAB75107.
DR
XX
XX Producing ubiquinone-10 for use as antioxidant, comprises using
PT microorganisms with reduced geranylgeranyl transferase and increased
PT decaprenyl diphosphate synthetase and hydroxybenzoic acid decaprenyl
PT transferase activity.
XX
XX Claim 8; Page 65-67; 82pp; Japanese.
XX
XX The present invention describes a method for producing ubiquinone-10 (1)
XX by culturing an microorganism that has: (a) reduced or no geranylgeranyl
XX transferase (crtE) activity; (b) strengthened decaprenyl diphosphate
XX synthase (DPPS) activity; or (c) strengthened hydroxybenzoic acid
XX decaprenyl transferase activity. (1) has cardiant and antioxidant
XX activities. The method is used for the production of ubiquinone-10,
XX which is useful for alleviating symptoms of heart disease and as a
XX antioxidant. The present sequence encodes a Rhodobacter sphaeroides
XX decaprenyl diphosphate synthase protein which is given in the present
XX invention.
XX
XX Sequence 999 BP; 164 A; 338 C; 351 G; 146 T; 0 other;
SQ

```

Query Match 68.8%; Score 17.2; DB 22; Length 999;
 Best Local Similarity 86.4%; Pred. No. 2.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ggcgcagcaggtgcatgcagc 22
   | | | | | | | | | | | | | | |
Db 954 GCGCAGCGGGTGTCTCGGCAGC 933

RESULT 35
AAQ31876/C
ID AAQ31876 standard; DNA; 1462 BP.
XX
XX AAQ31876:
AC
XX 22-APR-1993 (first entry)
XX
XX Cyclin D2 pseudogene.
DE
XX Cyclin; D2; D3; lambdaD2-G1; lambdaD3-G5; destruction box; A; B;
KM ubiquitin-dependant; degradation; similarity; D1; E; ss.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 378..1327
FT /*tag= a
FT 570..661
FT /*tag= b
FT /note= "Insertion"
FT 495..497
FT /*tag= c
FT

```


PF 23-OCT-1996; 96WO-US17327.
 XX
 PR 23-OCT-1995; 95US-0546934.
 XX
 PA (BETA-) BETAGENE INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Clark SA, Newgard CB, Normington KD, Thigpen AF;
 PI
 DR WPI: 1997-29737/27.
 DR P-PSDB; AAM17835.
 XX
 PT Use of glucose transporter, GLUT-2 or GLUT-2 chimera(s) for
 PT cell-killing - used in negative and double selection protocols and
 PT screening methods, for cancer treatment and treatment of diabetes
 XX
 PS Claim 138; Page 104-106, 169pp; English.
 XX
 CC This cDNA sequence codes for human glucose transporter GLUT-1
 CC (AAM17835). A claimed polynucleotide comprises a contiguous nucleic
 CC acid sequence from human GLUT-1 cDNA and rat GLUT-2 cDNA (see
 CC AAM17835). It encodes a GLUT-1/GLUT-2 chimeric transporter that
 CC confers glucose sensing capacity to a cell but which does not
 CC render the cell subject to diabetic immune destruction, and which
 CC does not transport streptozotocin. The claimed polynucleotide can
 CC be administered to a patient to treat diabetes, or expressed in
 CC a cell to prepare a recombinant cell that secretes insulin in
 CC response to glucose and which can be administered to a patient to
 CC treat diabetes.
 CC
 SQ Sequence 1815 BP; 324 A; 572 C; 520 G; 399 T; 0 other;

Query Match 68.8%; Score 17.2; DB 18; Length 1815;
 Best Local Similarity 86.4%; Pred. No. 2.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 cagcaggtcgcacgcacatt 25
 Db 470 CAGCAGCTTCATCATCAGCATT 449
 |||||

RESULT 38
 AAH68252/C
 ID AAH68252 standard; DNA; 1875 BP.
 XX
 AC AAH68252;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 3287.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 OS
 PN EPI108790-A2.
 PN
 PD 20-JUN-2001.
 PD
 PF 18-DEC-2000; 2000EP-0127688.
 PF
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 PI
 DR WPI: 2001-376931/40.

DR P-PSDB; AAG93033.
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 8; SEQ ID NO: 3287; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 1875 BP; 332 A; 647 C; 588 G; 308 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 1875;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgcacgcacgc 22
 Db 1803 GCCCAGCAGGAGCATCGCAATC 1782
 |||||

RESULT 39
 AAH56335
 ID AAH56335 standard; DNA; 1896 BP.
 XX
 AC AAH56335;
 XX
 DT 05-SEP-2001 (first entry)
 XX
 DE DNA encoding Chlamydia trachomatis PmpH(N-term) fusion protein.
 XX
 KW Chlamydia; vaccine; infection; fusion protein; antigen;
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
 KW acute respiratory tract infection; Cap1; CM529; OMCB;
 KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.
 XX
 OS Chlamydia trachomatis.
 OS
 PN WO200140474-A2.
 PN
 PD 07-JUN-2001.
 PD
 PF 04-DEC-2000; 2000WO-US32919.
 PF
 PR 03-DEC-1999; 99US-0454684.
 PR 19-APR-2000; 2000US-0556877.
 PR 20-JUN-2000; 2000US-0598419.
 XX
 PA (CORI-) CORIAX CORP.
 PA
 PI Probst P, Bhalla A, Skeiky YAW, Fling SP, Scholler J;
 PI
 DR WPI: 2001-374831/39.
 DR
 PT Chlamydia polypeptides and fusion proteins useful for preventing pelvic
 PT inflammatory disease, trachoma, acute respiratory tract infections,
 PT atherosclerosis and heart disease -
 PT
 XX
 XX Example 10; Page 274; 295pp; English.

XX The present nucleotide sequence is provided in a specification
 CC relating to compounds and methods for the treatment and diagnosis of
 CC Chlamydia infection. The compounds provided include polypeptides and
 CC fusion proteins comprising immunogenic portions of Chlamydia antigens
 CC and DNA sequences encoding such polypeptides. They are useful for
 CC vaccinating against chlamydial infection, which causes pelvic
 CC inflammatory disease, trachoma, acute respiratory tract infections,
 CC atherosclerosis and heart disease.

XX Sequence 1896 BP; 478 A; 425 C; 484 G; 509 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 1896;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcagcaggtgcacgcagcat 24
 ||||| ||||| |||||
 Db 1810 gcagcgggtgcacgcacacat 1831

RESULT 40

AAH56338
 ID AAH56338 standard; DNA: 2148 BP.

AC AAH56338;

DT 05-SEP-2001 (first entry)

DE DNA encoding Chlamydia trachomatis pmpH(C-term) fusion protein.

XX Chlamydia; vaccine; infection; fusion protein; antigen;

KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;

KM acute respiratory tract infection; CapI; CT529; OMCB;

KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.

OS Chlamydia trachomatis.

PN W0200140474-A2.

PD 07-JUN-2001.

PF 04-DEC-2000; 2000WO-US32919.

PR 03-DEC-1999; 99US-0454684.

PR 19-APR-2000; 2000US-0556877.

PR 20-JUN-2000; 2000US-0598419.

XX (CORI-) CORIXA CORP.

PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;

XX WPI: 2001-374831/39.

DR Chlamydia polypeptides and fusion proteins useful for preventing pelvic

XX inflammatory disease, trachoma, acute respiratory tract infections,

XX atherosclerosis and heart disease -

XX Example 10; Page 276-277; 295pp; English.

XX The present nucleotide sequence is provided in a specification

XX relating to compounds and methods for the treatment and diagnosis of

XX chlamydial infection. The compounds provided include polypeptides and

XX fusion proteins comprising immunogenic portions of Chlamydia antigens

XX and DNA sequences encoding such polypeptides. They are useful for

XX vaccinating against chlamydial infection, which causes pelvic

XX inflammatory disease, trachoma, acute respiratory tract infections,

XX atherosclerosis and heart disease.

XX Sequence 2148 BP; 544 A; 494 C; 528 G; 582 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 2148;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcagcaggtgcacgcagcat 24
 ||||| ||||| |||||
 Db 466 gcagcgggtgcacgcacacat 487

RESULT 41

AAH65400/C
 ID AAH65400 standard; DNA: 2265 BP.

AC AAH65400;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 435.

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

OS EPI108790-A2.

PN 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

PA (KYOWA) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR P-PSDB: MAG90181.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 435; 246bp + Sequence listing; English.

XX The present invention provides a number of nucleotide and protein

XX sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These

XX are useful for identifying the mutation point of a gene derived from a

XX mutant of Corynebacterium bacteria, measuring expression amount and

XX analysing the expression profile or expression pattern of a gene derived

XX from Corynebacterium bacteria, and identifying a homologue of a gene derived

XX from Corynebacterium bacteria. Corynebacterium bacteria are useful for producing

XX amino acids, nucleic acids, vitamins, saccharides and organic acids,

XX particularly L-lysine. The present sequence is a nucleic acid described

XX in the exemplification of the invention.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from the

XX European Patent Office.

XX Sequence 2265 BP; 513 A; 769 C; 556 G; 427 T; 0 other;

XX Query Match 68.8%; Score 17.2; DB 22; Length 2265;

XX Best Local Similarity 86.4%; Pred. No. 2.9e+02;

XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcagcaggtgcacgcagcat 24
 ||||| ||||| |||||
 Db 637 GCAGCAGGTGCATCTCATCTT 616

```

RESULT 42
AA224632/C
ID AA224632 standard; cDNA; 2850 BP.
XX
AC AA224632;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human lung tumor associated polynucleotide.
XX
KW Human; lung tumor; lung cancer; T cell stimulation; ss.
XX
OS Homo sapiens.
XX
PN W09947674-A2.
XX
PD 23-SEP-1999.
XX
PF 17-MAR-1999; 99WO-US05798.
XX
PR 18-MAR-1998; 98US-0040802.
PR 18-MAR-1998; 98US-0040984.
PR 27-JUL-1998; 98US-0123912.
PR 27-JUL-1998; 98US-0123933.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Wang T;
XX
DR WPI: 1999-571839/48.
XX
PT New isolated lung tumor polynucleotides, used to develop products for
PT the treatment, prevention and monitoring the progression of lung cancer
PI
PS Claim 12; Page 123-124; 148pp; English.
XX
CC The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.
CC
SO Sequence 2850 BP; 602 A; 800 C; 753 G; 695 T; 0 other;

Query Match 68.8%; Score 17.2; DB 20; Length 2850;
Best Local Similarity 86.4%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 cagcaggtgcatcgccagcatt 25
    ||||| |||| |||||
DB 485 CAGCAGTTCATCATCAGCATT 464

RESULT 43
AAC65871/C
ID AAC65871 standard; cDNA; 2856 BP.
XX
AC AAC65871;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated cDNA for config 33.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection; ss.
XX

```

```

OS Homo sapiens.
XX
PN W0200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L;
XX
DR WPI: 2000-628399/50.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient.
XX
PS Claim 25a; Page 156-157; 261pp; English.
XX
CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.
XX
SO Sequence 2856 BP; 602 A; 804 C; 753 G; 697 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 2856;
Best Local Similarity 86.4%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 cagcaggtgcatcgccagcatt 25
    ||||| |||| |||||
DB 485 CAGCAGTTCATCATCAGCATT 464

RESULT 44
AAF98714/C
ID AAF98714 standard; DNA; 2893 BP.
XX
AC AAF98714;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human late stage ovarian tumour polynucleotide marker 15.
XX
KW Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200118542-A2.
XX
PD 15-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US24199.
XX
PR 03-SEP-1999; 99US-0152547.
PR 16-MAR-2000; 2000US-0190347.
XX

```

XX 21-MAR-2000: 2000US-0191321.
 PT 31-MAY-2000: 2000US-0208382.
 PR 20-JUL-2000: 2000US-0220467.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PI Lee J, Thompson P, Little J;
 DR WPI: 2001-211428/21.
 XX
 PT Detection, assessment, prevention and therapy of ovarian cancer,
 PT comprises detecting changes in the expression of a variety of markers -
 XX
 PS Claim 1: Page 1180-1181; 1198pp: English.
 XX
 CC The present invention describes a method for assessing whether a patient
 CC is afflicted with ovarian cancer by comparing: (1) the expression of a
 CC marker (1) (see AAF98594 to AAF98730), in a patient sample; and (2) the
 CC normal level of expression of (1) in a control non-ovarian cancer
 CC sample, where a significant difference between the level of expression
 CC in (a) and (b) is an indication that the patient is afflicted with
 CC ovarian cancer. (1) have cytostatic activities and can be used in
 CC antisense gene therapy. The method, compositions and kits from the
 CC present invention can be used for: (1) assessing and treating ovarian
 CC cancer; (2) making isolated hybridoma, which produces an antibody useful
 CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
 CC patient. AAF98573 to AAF98593 represent human kinase marker primers and
 CC probes which are used in the exemplification of the present invention.
 XX
 SO Sequence 2893 BP: 607 A: 822 C: 765 G: 699 T: 0 other:

Query Match 68.8%; Score 17.2; DB 22; Length 2893;
 Best Local Similarity 86.4%; Pred. NO. 3e+02; 3; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 4 cagcagtgatcgcgcagcat 25
 ||||||| ||||| |||||||
 DB 522 CAGCAGCTTCATCAGCAGCAT 501

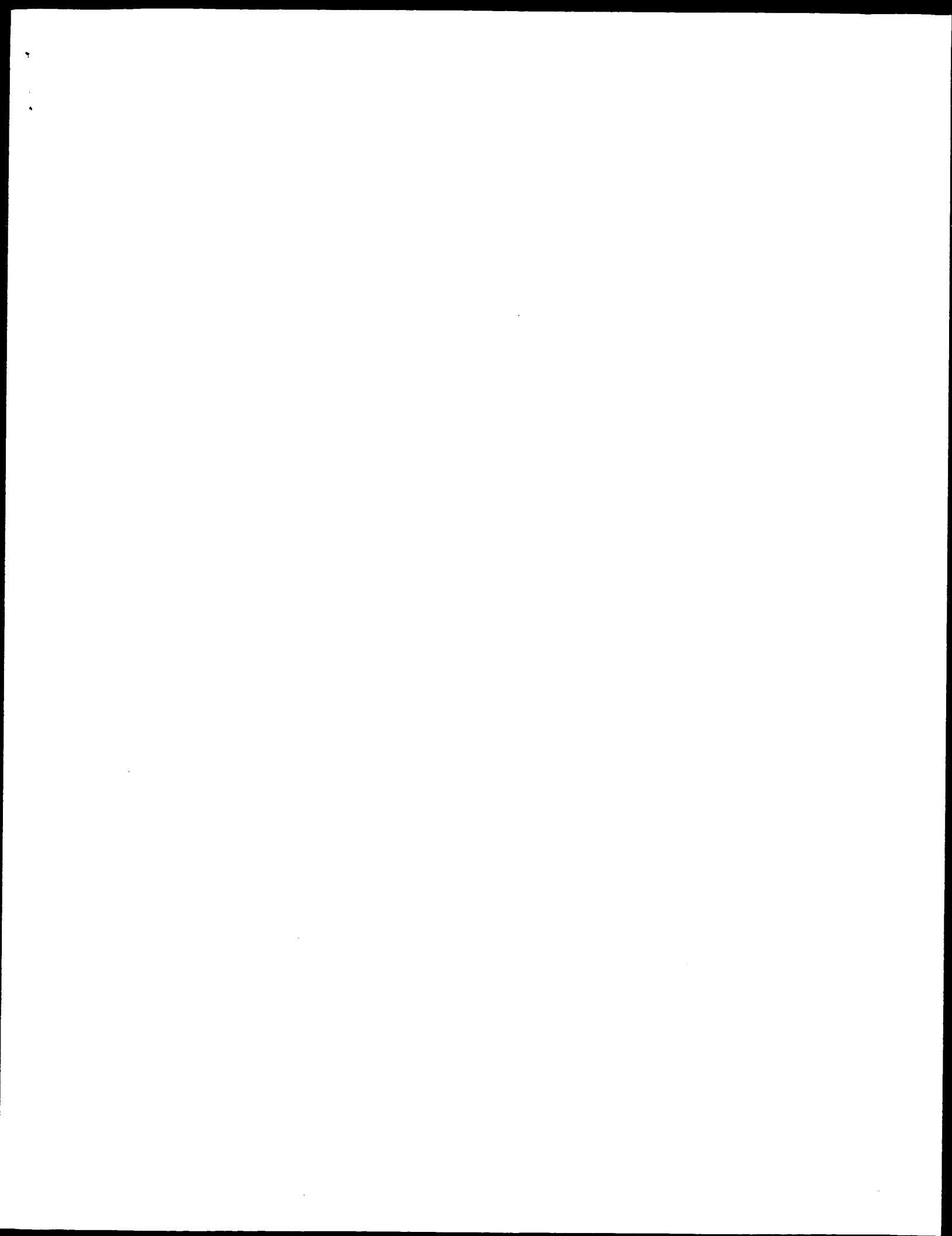
RESULT 45
 ID AAA81538 standard: DNA: 5532 BP.
 XX
 AC AAA81538;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_85 SEQ ID NO:85.
 XX
 KW Neisseria meningitidis: Neisseria gonorrhoeae: genome: immunogenic;
 KW antigen: vaccine: diagnosis: infection: antibacterial: identification;
 KW Meningococcus B; Menb; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999: 99WO-US23573.
 XX
 PR 09-OCT-1998: 98US-0103794.
 PR 30-APR-1999: 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tetelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratli G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI: 2000-318079/27.

XX
 PT isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7: Page 1480-1482; 1760pp: English.
 XX

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis Menb polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX

SO Sequence 5532 BP: 1241 A: 1503 C: 1452 G: 1336 T: 0 other:
 Query Match 68.8%; Score 17.2; DB 21; Length 5532;
 Best Local Similarity 86.4%; Pred. NO. 3.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 3 gcagcagtgatcgcgcagcat 24
 ||||||| ||||| |||||||
 DB 5390 gcagcagggctcgcgcagcat 5411

Search completed: December 26, 2001, 12:18:36
 Job time: 6877 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:23:07 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 ggcagcagtgatcgcagcagcat 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.6	70.4	2126	2	US-08-484-397A-22
6	17.6	70.4	2126	2	US-08-789-354-1
7	17.6	70.4	2126	3	US-09-110-937-1
8	17.6	70.4	2126	3	US-09-058-725B-1
9	17.6	70.4	2126	3	US-09-232-857-1
10	17.2	68.8	423	1	US-08-470-179-159
11	17.2	68.8	1462	2	US-08-464-517-30
12	17.2	68.8	1462	2	US-08-246-361A-30
13	17.2	68.8	1462	2	US-08-463-772-30
14	17.2	68.8	1462	5	PCR-US93-05000-30
15	16.8	67.2	9551	1	US-08-056-200-93
16	16.8	67.2	9551	2	US-08-800-644-93
17	16.6	66.4	414	3	US-08-804-180C-1
18	16.6	66.4	423	1	US-08-470-179-45
19	16.6	66.4	423	1	US-08-470-179-47
20	16.6	66.4	484	6	5496550-5
21	16.6	66.4	561	1	US-08-067-684-13
22	16.6	66.4	561	1	US-08-008-898-13
23	16.6	66.4	561	2	US-08-459-818-13
24	16.6	66.4	561	2	US-08-889-666-13
25	16.6	66.4	561	2	US-08-465-078-13
26	16.6	66.4	561	2	US-08-725-776-13
27	16.6	66.4	561	2	US-08-488-062-13

28	16.6	66.4	561	3	US-08-228-208A-13	Sequence 13, Appl
29	16.6	66.4	561	5	PCT-US95-06726-35	Sequence 35, Appl
30	16.6	66.4	1289	2	US-08-344-833-1	Sequence 1, Appl
31	16.6	66.4	2188	1	US-07-706-872-2	Sequence 2, Appl
32	16.6	66.4	3223	2	US-08-620-694A-9	Sequence 9, Appl
33	16.6	66.4	3223	3	US-09-022-255-9	Sequence 9, Appl
34	16.6	66.4	3223	3	US-09-022-255-9	Sequence 9, Appl
35	16.6	66.4	3223	3	US-08-978-773-3	Sequence 9, Appl
36	16.6	66.4	3223	3	US-09-022-253-9	Sequence 9, Appl
37	16.6	66.4	3223	3	US-09-022-250-9	Sequence 9, Appl
38	16.6	66.4	3223	4	US-09-022-259-9	Sequence 9, Appl
39	16.6	66.4	3223	4	US-09-022-257-9	Sequence 9, Appl
40	16.6	66.4	3958	5	US-08-435-933-5	Sequence 5, Appl
41	16.6	66.4	3958	5	PCR-US96-06035-5	Sequence 5, Appl
42	16.6	66.4	8931	3	US-09-028-934-28	Sequence 28, Appl
43	16.6	66.4	4411529	4	US-09-103-840A-1	Sequence 1, Appl
44	16.2	64.8	117	6	5273901-8	Patent No. 5273901
45	16.2	64.8	117	6	5482709-7	Patent No. 5482709
46	16.2	64.8	423	1	US-08-470-179-144	Sequence 144, App
47	16.2	64.8	946	2	US-08-408-095-26	Sequence 26, Appl
48	16.2	64.0	423	1	US-08-470-179-75	Sequence 75, Appl
49	16	64.0	424	1	US-08-470-179-79	Sequence 79, Appl
50	16	64.0	588	4	US-09-475-316A-105	Sequence 105, App
51	16	64.0	1386	2	US-08-756-506-3	Sequence 3, Appl
52	16	64.0	1387	6	5270178-1	Patent No. 5270178
53	16	64.0	1755	6	5225537-1	Patent No. 5225537
54	16	64.0	2391	2	US-08-817-900-1	Sequence 1, Appl
55	16	64.0	2391	4	US-09-236-645-1	Sequence 1, Appl
56	16	64.0	11219	1	US-07-642-73AC-1	Sequence 1, Appl
57	16	64.0	11219	3	US-08-439-009A-1	Sequence 1, Appl
58	16	64.0	11725	2	US-08-756-506-1	Sequence 1, Appl
59	16	64.0	50341	1	US-08-247-901C-1	Sequence 1, Appl
60	16	64.0	50341	2	US-09-075-936-1	Sequence 1, Appl
61	16	64.0	52297	4	US-09-426-434-1	Sequence 1, Appl
62	16	64.0	52297	4	US-08-705-551-1	Sequence 1, Appl
63	15.8	63.2	643	4	US-08-861-774E-53	Sequence 53, Appl
64	15.8	63.2	4689	4	US-09-105-537-34	Sequence 34, Appl
65	15.8	63.2	5362	2	US-08-853-310-3	Sequence 3, Appl
66	15.8	63.2	11220	4	US-09-105-537-32	Sequence 32, Appl
67	15.8	63.2	13842	4	US-09-105-537-30	Sequence 30, Appl
68	15.8	63.2	36788	4	US-09-105-537-5	Sequence 5, Appl
69	15.8	63.2	36788	3	US-09-320-878-19	Sequence 19, Appl
70	15.6	62.4	57	2	US-07-814-220-26	Sequence 26, Appl
71	15.6	62.4	57	2	US-07-814-221-26	Sequence 26, Appl
72	15.6	62.4	62	2	US-07-814-220-25	Sequence 25, Appl
73	15.6	62.4	62	2	US-07-814-221-25	Sequence 25, Appl
74	15.6	62.4	93	4	US-08-556-978B-25	Sequence 25, Appl
75	15.6	62.4	93	4	US-08-556-978B-25	Sequence 25, Appl
76	15.6	62.4	303	4	US-08-556-978B-80	Sequence 80, Appl
77	15.6	62.4	303	4	US-08-556-978B-81	Sequence 81, Appl
78	15.6	62.4	336	2	US-07-814-220-3	Sequence 3, Appl
79	15.6	62.4	336	2	US-07-814-220-3	Sequence 4, Appl
80	15.6	62.4	336	2	US-07-814-221-3	Sequence 4, Appl
81	15.6	62.4	336	2	US-07-814-221-3	Sequence 4, Appl
82	15.6	62.4	423	1	US-08-470-179-157	Sequence 157, App
83	15.6	62.4	423	1	US-08-470-179-158	Sequence 158, App
84	15.6	62.4	588	2	US-08-924-695A-1	Sequence 1, Appl
85	15.6	62.4	588	2	US-08-924-695A-1	Sequence 2, Appl
86	15.6	62.4	636	4	US-09-328-111-226	Sequence 226, App
87	15.6	62.4	710	4	US-08-998-416-263	Sequence 263, App
88	15.6	62.4	748	4	US-08-998-416-263	Sequence 263, App
89	15.6	62.4	751	4	US-08-998-416-399	Sequence 399, App
90	15.6	62.4	954	3	US-08-911-853-24	Sequence 24, Appl
91	15.6	62.4	954	4	US-09-479-409-24	Sequence 24, Appl
92	15.6	62.4	1057	1	US-08-147-784-1	Sequence 1, Appl
93	15.6	62.4	1057	4	US-08-195-967-1	Sequence 1, Appl
94	15.6	62.4	1057	4	US-08-472-940-1	Sequence 1, Appl
95	15.6	62.4	1065	3	US-08-938-783-2	Sequence 2, Appl
96	15.6	62.4	1548	2	US-08-762-106-5	Sequence 5, Appl
97	15.6	62.4	1548	2	US-09-320-774-5	Sequence 5, Appl
98	15.6	62.4	1581	2	US-08-762-106-6	Sequence 6, Appl
99	15.6	62.4	1581	4	US-09-320-774-6	Sequence 6, Appl
100	15.6	62.4	1729	4	US-09-120-817-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagtgatgcgcacgatt 25
    |||
Db 79 gccgacagtgatgcgcacgatt 103

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagtgatgcgcacgatt 25
    |||
Db 79 gccgacagtgatgcgcacgatt 103

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBOS3A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: aminotransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
Query Match 100.0%: Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%: Pred. No. 0.048;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gccacgacggtgcatcgccagcatt 25
|||||
Db 195 gccacgacggtgcatcgccagcatt 219
RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC.number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC.number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biOD"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biOD RBS"
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 gccgcacagtgatccgcacatt 25
    |||
DB 195 gccgcacagtgatccgcacatt 219
```

```

RESULT 5
US-08-484-397A-22/c
; Sequence 22, Application US/08484397A
; Patent No. 5869055
```

```

GENERAL INFORMATION:
APPLICANT: Juan, Shao-Chieh
APPLICANT: Lichenstein, Henri S.
APPLICANT: Wright, Samuel D.
TITLE OF INVENTION: ANTI-INFLAMMATORY CD14 POLYPEPTIDES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: CA
COUNTRY: US
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,397A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Curry, Daniel R.
REGISTRATION NUMBER: 32,727
REFERENCE/DOCKET NUMBER: A-324A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805/447-1000
TELEFAX: 805/447-1090
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-484-397A-22
```

```

Query Match          70.4%; Score 17.6; DB 2; Length 40;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

QY 1 gccgcacagtgatccgcacatt 24
    ||| ||| ||| ||| ||| |||
DB 25 gccgcacagtgatccgcacatt 2
```

```

RESULT 6
US-08-789-354-1/c
; Sequence 1, Application US/08789354
; Patent No. 5851798
GENERAL INFORMATION:
APPLICANT: Shabon, Usman
APPLICANT: Bergsma, Derk
TITLE OF INVENTION: Cloning of Human GPR14 Re
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,354
FILING DATE: 27-JAN-1997
```



```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: P50610
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-789-354-1

Query Match          70.4%; Score 17.6; DB 2; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgccacat 24
||||| ||| ||| ||| ||| ||| |||
DB 1269 GCGCACGACGCCGATGCCACGAT 1246

RESULT 7
US-09-110-937-1/c
Sequence 1, Application US/09110937A
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: BERGSMAN, DEBK
TITLE OF INVENTION: CLONING OF HUMAN GPR14 RECEPTOR
FILE REFERENCE: P50610-1
CURRENT APPLICATION NUMBER: US/09/110,937A
CURRENT FILING DATE: 1998-07-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 1
LENGTH: 2126
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-110-937-1

Query Match          70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgccacat 24
||||| ||| ||| ||| ||| ||| |||
DB 1269 GCGCACGACGCCGATGCCACGAT 1246

RESULT 8
US-09-058-725B-1/c
Sequence 1, Application US/09058725B
Patent No. 6133420
GENERAL INFORMATION:
APPLICANT: Ames, Robert
APPLICANT: Sarau, Henry
APPLICANT: Foley, James
APPLICANT: Chamber, Jon
TITLE OF INVENTION: A Method of Finding Antagonist
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,725B
FILING DATE: April 10, 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/789,354
FILING DATE: 27-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GP50005-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-058-725B-1

Query Match          70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgccacat 24
||||| ||| ||| ||| ||| ||| |||
DB 1269 GCGCACGACGCCGATGCCACGAT 1246

RESULT 9
US-09-232-857-1/c
Sequence 1, Application US/09232857
Patent No. 6159700
GENERAL INFORMATION:
APPLICANT: DOUGLAS, STEPHEN
APPLICANT: WILLETTTE, ROBERT
APPLICANT: AYAR, NAMBI
APPLICANT: ROMANIC, ANNE
APPLICANT: KHANDOUDI, NASSIRAH
APPLICANT: GOUT, BERNARD
APPLICANT: AL-BARAZANJI, KAMAL
APPLICANT: AMES, ROBERT S.
APPLICANT: FOLEY, JAMES J.
APPLICANT: SARAU, HENRY
APPLICANT: CHAMBERS, JON K.
APPLICANT: SHABON, USMAN
TITLE OF INVENTION: A METHOD OF FINDING AGONIST
TITLE OF INVENTION: AND ANTAGONIST TO HUMAN AND RAT GPR14
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/232.857
FILING DATE: 15-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/789,354
FILING DATE: 27-JAN-1997
APPLICATION NUMBER: 60/074,075
FILING DATE: 09-FEB-1998
APPLICATION NUMBER: 09/058,725
FILING DATE: 10-APR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-50005-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-232-857-1

Query Match          70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcacgat 24
    ||||| ||| ||| |||||
Db 1269 GCGCAGCAGCGCATGCGCAGCAT 1246

RESULT 10
US-08-470-179-159/c
Sequence 159, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
APPLICANT: Huang Ph.D, Wei Mun
TITLE OF INVENTION: Method and Compositions for
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESS: Trask, Britt and Rossa
STREET: P.O. Box 2550
CITY: Salt Lake City
STATE: Utah
COUNTRY: USA
ZIP: 84110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,179
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sweigert Ph.D, Susan E.
REGISTRATION NUMBER: 36,289
REFERENCE/DOCKET NUMBER: 2601
TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 159:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: No. 5645994ardia farcinica
US-08-470-179-159

Query Match          68.8%; Score 17.2; DB 1; Length 423;
Best Local Similarity 86.4%; Pred. No. 68;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcacgc 22
    ||||| ||| ||| |||||
Db 294 GCGCAGCATCTCCATCGCCAGC 273

RESULT 11
US-08-464-517-30/c
Sequence 30, Application US/08464517
Patent No. 5869640
GENERAL INFORMATION:
APPLICANT: BEACH, David H.
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,517
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/963,308
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/888,178
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/701,514
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mathew P. Vincent
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-004C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
```


;;
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Milltia Drive
;; CITY: Lexington
;; STATE: Massachusetts
;; COUNTRY: US
;; ZIP: 02173
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/05000
;; FILING DATE: 19930525
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/888,178
;; FILING DATE: 26-MAY-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: CSHL91-02A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-861-6240
;; TELEFAX: 616-861-9540
;; INFORMATION FOR SEQ ID NO: 30:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1462 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; PCT-US93-05000-30
;
Query Match 68.8%; Score 17.2; DB 5; Length 1462;
Best Local Similarity 86.4%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
QY 1 ggcgcagcaggtgcacgcacg 22
|||||
Db 392 GCCCAGCAGCTGCATGCCACG 371
;
RESULT 15
US-08-056-200-93
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993

;;
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fedrick, Michael F.
;; REGISTRATION NUMBER: 36,799
;; REFERENCE/DOCKET NUMBER: NIH054.001A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (714) 760-0404
;; TELEFAX: (714) 760-9502
;; INFORMATION FOR SEQ ID NO: 93:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 9551 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1507..1644
;; FEATURE:
;; NAME/KEY: Intron
;; LOCATION: 1645..2511
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2512..8070
;; US-08-056-200-93

Query Match 67.2%; Score 16.8; DB 1; Length 9551;
Best Local Similarity 90.0%; Pred. No. 13e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
QY 4 cagcaggtgcacgcacgca 23
|||||
Db 6805 CAGCAGCTGCACGCCACGCA 6824

RESULT 16
US-08-800-644-93
; Sequence 93, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799

```

REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: NO
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Inton
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-800-644-93

```

```

Query Match      67.2%; Score 16.8; DB 2; Length 9551;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 cagcagtgatcgccagca 23
DB 6805 CAGCAGCTGCACCGCAGCA 6824

```

```

RESULT 17
US-08-804-180C-1
Sequence 1, Application US/08804180C
Patent No. 6107056
GENERAL INFORMATION:
APPLICANT: Martin K. Oaks
TITLE OF INVENTION: scTla-4 and Its Soluble Products
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas M. Wozny
STREET: 100 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 inch Disk
COMPUTER: IBM
OPERATING SYSTEM: DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,180C
FILING DATE: February 20, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Thomas M. Wozny
REGISTRATION NUMBER: 28,922
REFERENCE/DOCKET NUMBER: 3284-00003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 271-7590
TELEFAX: (414) 271-5770
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 414
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: no

```

```

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapien
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: lymphnode
IMMEDIATE SOURCE:
CLONE: 204/bsCTLA-4/PCR3
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 2q33
FEATURE:
NAME/KEY: Human scTla-4 gene
IDENTIFICATION METHOD: Found by experiment
OTHER INFORMATION: Expresses b7 binding protein
US-08-804-180C-1

```

```

Query Match      66.4%; Score 16.6; DB 3; Length 414;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 gcagcagtgatcgccagcatt 25
DB 35 GCAGCGAGCATCGCCAGCTTT 57

```

```

RESULT 18
US-08-470-179-45/C
Sequence 45, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
APPLICANT: Huang Ph.D. Mai Mun
TITLE OF INVENTION: Method and Compositions for
IDENTIFICATION OF SPECIES IN A SAMPLE
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: Trask, Britt and Rossa
STREET: P.O. Box 2550
CITY: Salt Lake City
STATE: Utah
COUNTRY: USA
ZIP: 84110
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,179
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sweigert Ph.D. Susan E.
REGISTRATION NUMBER: 36,289
REFERENCE/DOCKET NUMBER: 2601
TELECOMMUNICATION INFORMATION:
TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "gyra gene segment"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Borrelia coriaceae
US-08-470-179-45

```

```

Query Match      66.4%; Score 16.6; DB 1; Length 423;

```


Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatgcgcagcatt 25
||||| | ||||| ||||| ||
Db 38 GCAGCCGAGGCATGCCGCGCTT 60

RESULT 22

US-08-008-898-13
; Sequence 13, Application US/08008898
; Patent No. 5770197
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S
; APPLICANT: Ledbetter, Jeffrey A
; APPLICANT: Damle, Nitin K
; APPLICANT: Brady, William
; TITLE OF INVENTION: CTLA4 RECEPTOR AND METHODS FOR ITS USE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 201 South Lake Avenue, Suite 800
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/008.898
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/723,617
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, Saralynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 7848
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
US-08-008-898-13

Query Match

Best Local Similarity 66.4%; Score 16.6; DB 1; Length 561;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatgcgcagcatt 25
||||| | ||||| ||||| ||
Db 38 GCAGCCGAGGCATGCCGCGCTT 60

RESULT 23

US-08-459-818-13

; Sequence 13, Application US/08459818

; Patent No. 5851795

; GENERAL INFORMATION:

; APPLICANT: Linsley, Peter S.

; APPLICANT: Ledbetter, Jeffrey A.

; APPLICANT: Damle, Nitin K.

; APPLICANT: Brady, William

; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant & Gould

; STREET: 11150 Santa Monica Blvd., Suite 400

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90025

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: FastSeq 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/459,818

; FILING DATE: 02-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Adriano, Sarah B.

; REGISTRATION NUMBER: 34,470

; REFERENCE/DOCKET NUMBER: 30436.350502

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 310-445-1140

; TELEFAX: 310-445-9031

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 561 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..561

US-08-459-818-13

Query Match

Best Local Similarity 66.4%; Score 16.6; DB 2; Length 561;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatgcgcagcatt 25
||||| | ||||| ||||| ||
Db 38 GCAGCCGAGGCATGCCGCGCTT 60

RESULT 24

US-08-889-666-13

; Sequence 13, Application US/08889666

; Patent No. 5885579

; GENERAL INFORMATION:

; APPLICANT: Linsley, Peter S.

; APPLICANT: Ledbetter, Jeffrey A.

; APPLICANT: Damle, Nitin K.

; APPLICANT: Brady, William

; APPLICANT: Kiener, Peter A.

; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant & Gould

; STREET: 11150 Santa Monica Blvd., Suite 400

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90025

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/889,666
: FILING DATE: 08-JUL-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/375390
: FILING DATE: 18-JAN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Adriano, Sarah B.
: REGISTRATION NUMBER: 34,470
: REFERENCE/DOCKET NUMBER: 30436-35US01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 310-445-1140
: TELEFAX: 310-445-9031
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 561 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..561
: US-08-889-666-13

Query Match      66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 gcagcaggtgcacgcacgatt 25
   ||||| | ||||| ||||| ||
Db 38 GCAGCCGAGCATGCCACGCTTT 60

RESULT 25
US-08-465-078-13
: Sequence 13, Application US/08465078
: Patent No. 5885796
: GENERAL INFORMATION:
: APPLICANT: Linsley, Peter S.
: APPLICANT: Ledbetter, Jeffrey A.
: APPLICANT: Dame, Nitin K.
: APPLICANT: Brady, William
: APPLICANT: Kiener, Peter A.
: TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merchant & Gould
: STREET: 11150 Santa Monica Blvd., Suite 400
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90025
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/465,078
: FILING DATE: 05-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/375390
: FILING DATE: 18-JAN-1995
```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Adriano, Sarah B.
: REGISTRATION NUMBER: 34,470
: REFERENCE/DOCKET NUMBER: 30436-35US01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 310-445-1140
: TELEFAX: 310-445-9031
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 561 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..561
: US-08-465-078-13

Query Match      66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 gcagcaggtgcacgcacgatt 25
   ||||| | ||||| ||||| ||
Db 38 GCAGCCGAGCATGCCACGCTTT 60

RESULT 26
US-08-725-776-13
: Sequence 13, Application US/08725776
: Patent No. 5968510
: GENERAL INFORMATION:
: APPLICANT: Linsley, Peter S.
: APPLICANT: Ledbetter, Jeffrey A.
: APPLICANT: Dame, Nitin K.
: APPLICANT: Brady, William
: APPLICANT: Kiener, Peter A.
: TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merchant & Gould
: STREET: 11150 Santa Monica Blvd., Suite 400
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90025
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/725,776
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/375390
: FILING DATE: 18-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Adriano, Sarah B.
: REGISTRATION NUMBER: 34,470
: REFERENCE/DOCKET NUMBER: 30436-35US01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 310-445-1140
: TELEFAX: 310-445-9031
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 561 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
```


MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-725-776-13

Query Match 66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgcacgacatt 25
||||| 1 ||||| 111111111111
DB 38 GCAGCGGAGGATCGCCAGCTTT 60

RESULT 27
US-08-488-062-13
Sequence 13, Application US/08488062
Patent No. 5977318
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 1150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488.062
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-9031
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-488-062-13

Query Match 66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 gcacgagtgatcgcacgacatt 25
||||| 1 ||||| 111111111111

DB 38 GCAGCGGAGGATCGCCAGCTTT 60

RESULT 28
US-08-228-208A-13
Sequence 13, Application US/08228208A
Patent No. 6090914
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Wallace, Phillip M.
TITLE OF INVENTION: CTLA4/CD28lg HYBRID FUSION
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 1150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/228, 208A
FILING DATE: 15-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/008,898
FILING DATE: 22-JAN-1993
APPLICATION NUMBER: 07/723,617
FILING DATE: 27-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-30US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-228-208A-13

Query Match 66.4%; Score 16.6; DB 3; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 gcacgagtgatcgcacgacatt 25
||||| 1 ||||| 111111111111
DB 38 GCAGCGGAGGATCGCCAGCTTT 60

RESULT 29
PCT-US95-06726-35
Sequence 35, Application PC/TUS9506726
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Ligands for Induction of Antigen Specific Apoptosis 1

```

NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, suite 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06726
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/253,783
FILING DATE: 03 JUNE 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandagouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPI-016PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
PCT-US95-06726-35
Query Match 66.4%; Score 16.6; DB 5; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 gcagcagtgatccagccagcatt 25
||||| | ||||| |||
Db 38 GCAGCCGAGGCATGCCAGCTTT 60

```

```

APPLICATION NUMBER: US/08/344,833
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Blackstone, William B.
REGISTRATION NUMBER: 29,772
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bovine herpesvirus
FEATURE:
NAME/KEY: CDS
LOCATION: 142..1095
OTHER INFORMATION: /product= "Protein"
OTHER INFORMATION: /standard_name= "ORF-1"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 1..3
OTHER INFORMATION: /function= "stop-codon giv"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 4..141
OTHER INFORMATION: /function= "untranslated"
OTHER INFORMATION: "region"
US-08-344-833-1
Query Match 66.4%; Score 16.6; DB 2; Length 1289;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gcgcagcagtgatccagcaga 23
||||| | || | |||||
Db 1077 GCGCAGCAGCAGCAGCAGCAGCA 1055

```

```

RESULT 30
US-08-344-833-1/c
; Sequence 1, Application US/08344833
; Patent No. 5874280
GENERAL INFORMATION:
APPLICANT: Kell, G nther
TITLE OF INVENTION: Recombinant Bovine Herpesvirus
TITLE OF INVENTION: vaccine
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

RESULT 31
US-07-706-872-2/c
; Sequence 2, Application US/07706872
; Patent No. 5237056
GENERAL INFORMATION:
APPLICANT: Fischbach, Gerald D.
TITLE OF INVENTION: Identification of a Protein Which
TITLE OF INVENTION: Promotes the Synthesis of Acetylcholine Receptors and Uses
TITLE OF INVENTION: Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706,872
FILING DATE: 19910529
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia

```



```

? ORGANISM: Human
? STRAIN: IL-17 R (hCTLA8 receptor)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 93..2693
US-09-022-255-9

```

Query Match Similarity	66.4%	Score	16.6	DB	3	Length	3223
Best Local Similarity	83.6%	Pred.	No.1.4e+02				
Matches	19	Conservative	0	Mismatches	4	Indels	0
						Caps	0

```

RESULT 34
US-09-022-696--9/c
: Sequence 9, Application US/09022696
: Patent No. 6072037
: GENERAL INFORMATION:
: APPLICANT: Yao, Zhengbin
: APPLICANT: Springs, Melanie
: APPLICANT: Fasnlow, William
: TITLE OF INVENTION: No. 6072037el Receptor That Binds IL-17
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Immunex Corporation
: STREET: 51 University Street
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98101
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple Power Macintosh
: OPERATING SYSTEM: Apple Operating System 7.5.5
: SOFTWARE: Microsoft Word for Apple, Version 6.0.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/022,696
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/620,694
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: USSN 08/410,535
: FILING DATE: 23 MARCH 1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Perkins, Patricia Anne
: REGISTRATION NUMBER: 34,695
: REFERENCE/DOCKET NUMBER: 2617-B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206)587-0430
: TELEFAX: (206)
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3223 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Human
: STRAIN: IL-17 R (NCTLA8 receptor)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 93..2693

```

US-09-022-696-9

Query Match	66.48;	Score 16.6;	DB 3:	Length 3223;
Best local Similarity	82.6%;	Pred. No. 1.4e+02;		
Matches 19;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;

```

RESULT    35 US-08-978-773-3/C
; Sequence 3, Application US/08978773
; Patent No. 6083906
; GENERAL INFORMATION:
; APPLICANT: Trout, Anthony
; TITLE OF INVENTION: Method of Regulating Nitric Oxide Production
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple PowerMacintosh
; OPERATING SYSTEM: Apple Operating System 7.5.5
; SOFTWARE: Microsoft Word for PowerMacintosh, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,773
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/052,525
FILING DATE: 27 NOVEMBER 1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2623-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ. ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human
IMMEDIATE SOURCE:
CLONE: IL-17R
FEATURE:
NAME/KEY: CDS
LOCATION: 93..2690
US-08-978-773-3
Query Match      66.4%; Score 16.6; DB 3; Length 3223;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0
OY          I ggcgacgcagctcaccgccaga 23
             | ||||| ||| ||| |||||
Db          158 GACGAGCAGGACGAGCCCGACA 136

```

Query	1	gctgcagtcaggtgcacgcgcagca	23
Db	158	GAGCAGCAGCAGCAGCAGCCCGACGA	136
		Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
		Query March 66.4%; Score 16.6; DB 3; Length 3323;	
		Best Local Similarity 82.6%; Pred. No. 1.4e+02;	
		Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	


```

1  APPLICANT: Spriggs, Melanie
2  APPLICANT: Fanslow, William
3  TITLE OF INVENTION: No. 6191104e1 Receptor That Binds IL-17
4  NUMBER OF SEQUENCES: 10
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Immunex Corporation
7  STREET: 51 University Street
8  CITY: Seattle
9  STATE: WA
10 COUNTRY: USA
11 ZIP: 98101
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: Apple Power Macintosh
16 OPERATING SYSTEM: Apple Operating System 7.5.5
17 SOFTWARE: Microsoft Word for Apple, Version 6.0.1
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/09/022,259
21 FILING DATE:
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 08/620,694
25 FILING DATE:
26
27 CLASSIFICATION:
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: 08/620,694
30 FILING DATE:
31
32 CLASSIFICATION:
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: USSN 08/410,535
35 FILING DATE: 23 MARCH 1995
36 CLASSIFICATION:
37 ATTORNEY/AGENT INFORMATION:
38 NAME: Perkins, Patricia Anne
39 REGISTRATION NUMBER: 34,695
40 REFERENCE/DOCKET NUMBER: 2617-B
41 TELECOMMUNICATION INFORMATION:
42 TELEPHONE: (206)587-0430
43 TELEFAX: (206)
44
45 INFORMATION FOR SEQ. ID NO: 9:
46 SEQUENCE CHARACTERISTICS:
47 LENGTH: 3223 base pairs
48 TYPE: nucleic acid
49 STRANDEDNESS: single
50 TOPOLOGY: linear
51 MOLECULE TYPE: cDNA to mRNA
52 HYPOTHETICAL: NO
53 ANTI-SENSE: NO
54 ORIGINAL SOURCE:
55 ORGANISM: Human
56 STRAIN: IL-17 R (hCTL8 receptor)
57
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: 93..2693
61
62 US-09-022-259-9
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```

1  ADDRESSEE: Immunex Corporation
2  STREET: 51 University Street
3  CITY: Seattle
4  STATE: WA
5  COUNTRY: USA
6  ZIP: 98101
7
8  COMPUTER READABLE FORM:
9  MEDIUM TYPE: Floppy disk
10 COMPUTER: Apple Power Macintosh
11 OPERATING SYSTEM: Apple Operating System 7.5.5
12 SOFTWARE: Microsoft Word for Apple, Version 6.0.1
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/09/022,257
16 FILING DATE:
17 CLASSIFICATION:
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: 08/620,694
20 FILING DATE:
21 CLASSIFICATION:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: USN 08/410,535
24 FILING DATE: 23 MARCH 1995
25 CLASSIFICATION:
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Perkins, Patricia Anne
28 REGISTRATION NUMBER: 34,695
29 REFERENCE/DOCKET NUMBER: 2617-B
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (206)587-0430
32 TELEFAX: (206)
33
34 INFORMATION FOR SEQ. ID NO: 9:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 3223 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: cDNA to mRNA
41 HYPOTHETICAL: NO
42 ANTI-SENSE: NO
43 ORIGINAL SOURCE:
44 ORGANISM: Human
45 STRAIN: IL-17 R (hCTL8 receptor)
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 93..2693
49
50 US-09-022-257-9
51
52 Query Match 66.4%; Score 16.6; DB 4; Length 3223;
53 Best Local Similarity 82.6%; Pred. No. 1.4e+02;
54 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0.
55
56 QY 1 ggcgacgaagtcatccacga 23
57 | | | | | | | | | | | | | |
58 Db 158 GAGCAGCAGAGCAGCAGCCACGA 136
59
60 RESULT 40
61 US-08-435-933-5
62 ; Sequence 5, Application US/08435933
63 ; Patent No. 5693492
64 ; GENERAL INFORMATION:
65 ; APPLICANT: Cully, Doris F.
66 ; APPLICANT: Arena, Joseph P.
67 ; APPLICANT: Paresi, Philip S.
68 ; APPLICANT: Liu, Ken K.
69 ; TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
70 ; NUMBER OF SEQUENCES: 6
71 ; CORRESPONDENCE ADDRESS:
72 ; ADDRESSEE: John W. Wallen III
73 ; STREET: 126 East Lincoln Avenue
74 ; CITY: Rahway

```

STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,933
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Wallen, Iii John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-435-933-5

Query Match 66.4% Score 16.6; DB 1; Length 3958;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcgcagcaggtgcacgcagca 23
||||||| || |||||
Db 289 GCGCAGCAGCAGCAGCAGCA 311

RESULT 41
PCT-US96-06035-5
GENERAL INFORMATION:
SEQUENCE 5, Application PC/TUS9606035
APPLICANT: Cully, Doris F.
APPLICANT: Arena, Joseph P.
APPLICANT: Pares, Philip S.
APPLICANT: Liu, Ken K.
TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jody M. Gieser
STREET: 126 East Lincoln Avenue - P.O. BOX 2000-0907
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06035
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gieser, Jody M.
REGISTRATION NUMBER: 32,838
REFERENCE/DOCKET NUMBER: 19264 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3046
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 3958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US96-06035-5

Query Match 66.4% Score 16.6; DB 5; Length 3958;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcgcagcaggtgcacgcagca 23
||||||| || |||||
Db 289 GCGCAGCAGCAGCAGCAGCA 311

RESULT 42
US-09-028-934-28/c
SEQUENCE 28, Application US/09028934
PATENT NO. 6117670
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Lam, Steven T.
APPLICANT: Hammer, Philip E.
APPLICANT: van Pee, Karl-Heinz
APPLICANT: Kliner, Sabine
APPLICANT: Young, Thomas R.
TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6117670artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/028,934
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1506/C1P7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 8931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Burkholderia cepacia

```
FEATURE:
NAME/KEY: CDS
LOCATION: 657..2267
OTHER INFORMATION: /product= "PrnA"
FEATURE:
NAME/KEY: CDS
LOCATION: 2270..3355
OTHER INFORMATION: /product= "PrnB"
FEATURE:
NAME/KEY: CDS
LOCATION: 3421..5121
OTHER INFORMATION: /product= "PrnC"
FEATURE:
NAME/KEY: CDS
LOCATION: 5145..6266
OTHER INFORMATION: /product= "PrnD"
US-09-028-934-28
```

```
Query Match      66.4%; Score 16.6; DB 3; Length 8931;
Best Local Similarity 82.6%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 gccagcaggtgcatcgccagca 23
||||| ||| ||| ||| ||| |||
Db 5010 GTGACAGCAGCTTCATCGCGGCA 4988
```

```
RESULT 43
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match      66.4%; Score 16.6; DB 4; Length 4411529;
Best Local Similarity 86.4%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 ccgacaggtgcatcgccagca 23
||||| ||| ||| ||| ||| |||
Db 962021 ccgacagacccatcgccagca 962042
```

```
RESULT 44
5273901-8
Patent No. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON, SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
SPOROZOITE 21.5 KB ANTIGEN, AC-6B
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/581.693
FILING DATE: 12-SEP-1990
```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1988
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 8
LENGTH: 117
5273901-8
```

```
Query Match      64.8%; Score 16.2; DB 6; Length 117;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 gccagcaggtgcatcgccagca 23
||||| ||| ||| ||| ||| |||
Db 49 gccagcagcagcaccgcca 69
```

```
RESULT 45
5482709-7
Patent No. 5482709
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON, SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148.432
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 581,693
FILING DATE: 12-SEP-1990
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1989
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 7
LENGTH: 117
5482709-7
```

```
Query Match      64.8%; Score 16.2; DB 6; Length 117;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 gccagcaggtgcatcgccagca 23
||||| ||| ||| ||| ||| |||
Db 49 gccagcagcagcaccgcca 69
```

Search completed: December 26, 2001, 12:23:53
Job time: 7190 sec